1. An Introduction to the Human Body
    1. Introduction to Human Physiology
    2. Levels of Organization
    3. Functions of Human Life
    4. Requirements for Human Life
2. The Chemical Level of Organization
    1. Introduction
    2. Fundametals of Matter: Elements and Their Relationships
    3. Chemical Reactions
    4. Inorganic Compounds Essential to Human Functioning
    5. Organic Compounds Essential to Human Functioning
3. The Cellular Level of Organization
    1. Introduction to Cells
    2. The Plasma Membrane
    3. The Cytoplasm and Cellular Organelles
    4. Connections Between Cells and Cellular Activities
    5. Cell Growth and Division
    6. The Special Case of Sex Cells
    7. Cellular Differentiation
4. DNA and Protein Synthesis
    1. Determining the Structure of DNA
    2. DNA Replication
    3. Protein Synthesis
    4. Gene Regulation and Expression
    5. Biotechnology
5. Cellular Metabolism
    1. Energy and Metabolism
6. Cellular Respiration
    1. Overview of Cellular Respiration
    2. Glycolysis
    3. Oxidation of Pyruvate and the Citric Acid Cycle

    4. Oxidative Phosphorylation
    5. Metabolism without Oxygen
7. Membrane Transport
    1. Introduction to Membrane Transport
    2. Forces That Move Ions and Molecules
8. Cellular Communication
    1. Introduction to Cellular Communication
    2. Types of Signal Molecules
    3. Signal Molecule Release, Transport, and Binding
    4. Intracellular Signal Transduction

Introduction to Human Physiology

This chapter provides an introduction to human physiology. It is intended to help students understand key concepts that will help them navigate the specific details in later chapters.

The Study of Physiology

Physiology is the study of body function. Students taking physiology are generally expected to already know basic anatomy, which is the study of body structure and the identification and naming of the parts that make up the structure. In anatomy it was important to learn the basic function of individual parts to help you understand why those parts take the form they do. In physiology you will study the functions in more detail focusing on the interactions among the parts and the underlying chemistry that drives the physiology.

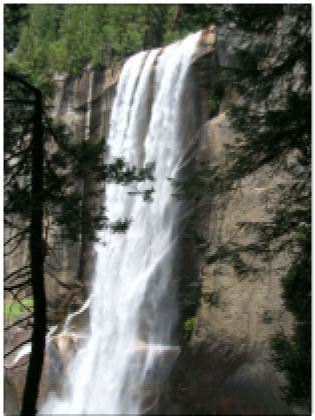
Function requires parts to act in a coordinated way. These coordinated activities are the processes that sustain the living state and allow individual cells, organs, and organisms to carry out specialized functions. Relationships and communication are not observable in the same way as anatomical parts. As a student new to physiology it is important that you begin by focusing on the process terminology and process order. Because these processes are generally dynamic and recursive (they loop back on themselves) it also is important that you work on understanding how changes ripple through a process when various parts change in a specific way.

There are too many possibilities to memorize all the processes and process states that might occur. Fortunately, there are a limited number of basic processes that seem to form the basis of most physiological activities. Understanding these underlying processes will help you make sense of the multitude of specialized processes you will encounter later. It is important that you focus on understanding the dynamics of these processes in the early chapters because they will be used over and over in the later chapters

Dynamic Kinetic Stability

You learned in chemistry that physical and chemical systems move toward the most stable, lowest energy, state. The progression follows the physical laws of thermodynamics and results in higher entropy, or disorder, with each physical or chemical interaction. The more disorder in the system the more stable its state. We stated in the previous section that the physiology of living things requires highly ordered interactions. From this perspective the existence of life would seem to violate the basic laws of physics and chemistry.

The chemist, Addy Pross, has pointed out that static stability following the rules of thermodynamics is not the only kind of stability found in nature. Another type of stability, dynamic kinetic stability (DKS), is also found in physical and chemical systems. Think of a waterfall. The waterfall appears to be a relatively stable structure. And yet the individual water molecules from which it is made are each in an unstable state as they flow downhill to their lowest stable point.



While it appears stable, the waterfall is made of a persistent flow of changing water molecules.

The stability of the waterfall as a physical structure is maintained by the persistent unstable (kinetic) state of the individual water molecules from which it is made. It should be obvious that a DKS system does not violate the laws of thermodynamics; however, to persist it does require a continual influx of outside energy and materials. There are many examples of non-living DKS systems, but what is important here is that all living biological cells, organisms, and populations are DKS systems. Their parts continually need to be replaced and repaired, which requires a constant influx of outside energy and materials. When the supply of outside energy or key materials is disrupted the DKS can no longer be maintained and the materials making up the living system revert to the static entropy rules and the state we call death.

Homeostasis

The key dogma of modern physiology is homeostasis, the ability of living organisms to maintain a relatively constant state. It should be obvious from the previous section that homeostasis can be viewed as the organism maintaining its dynamic kinetic stability. The focus of physiology is to understand how the organism accomplishes this.

Nothing in physiology happens by magic. We all have a tendency to get lazy and attribute conscious activity to cells or organs. Statements like: “The cell **determines** that the calcium levels are low,” or “The pituitary **tells** the adrenal cortex to release cortisol,” imply magical thinking. When you find yourself using terms like these as you try to explain a process or answer a question it is important that you stop and translate the magical term into a cause and effect process.

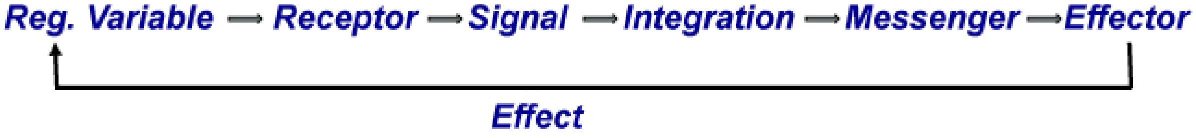
Cause and effect refers to a process where one action causes a second action to occur. **Order is important**, so when you are dealing with cause and effect you need to pay attention to the order in which the actions occur. In physiology cause and effect actions generally do not occur as a stand-alone pair but are chained together. This means that action-1 causes action-2, action-2 in turn causes action-3, action-3 causes action-4, and so on.

Action-1 triggers a cascade of actions each dependent on the previous action, so **again, order is important**.

It also is important to pay attention to how any cause and effect chain ends, and eventually that will take you back to the first action in the chain. In other words, the chain will feed back on itself and turn into a loop. A simple way to think of this is to determine what the last effect in the chain is on the stimulus that triggered the first cause in the chain. If the last effect decreases this stimulus we call it a negative feedback loop. If the last effect increases the stimulus we call it a positive feedback loop. **Note that positive and negative refer to the increase or decrease on the initial stimulus, not whether the products of the last effect or regulated variable increase or decrease**. If the first response is to a high level of the regulated variable, then a cause and effect chain that decreases that regulated variable's level will decrease the initial stimulus. If the first response is to a low level of the regulated variable, then a cause and effect chain that increases that regulated variable's level also will decrease the initial stimulus.

Negative Feedback Pathways

By reducing the conditions that led to the initial response (which may involve increasing or decreasing the regulated variable) negative feedback loops reduce the activity of the response. As the conditions of the regulated variable change the process slows and stops. This is the primary mechanism by which homeostasis is maintained without resorting to magic. When you are trying to understand a specific homeostatic (negative feedback) loop you should try to identify the following seven components (Figure 2). Learning and using this list of components can help you identify where you are lacking detail and need to do more research to increase your understanding.



Seven steps in a negative feedback loop.

1. The **Regulated Variable** is the condition being maintained within a narrow range known as the "set point." A regulated variable can be anything from the concentration of a specific chemical, to the speed of a vehicle, to the temperature of the blood in body.
2. The **Receptor** is the sensor that reacts with the specific regulated variable and continuously transduces (translates) the variable's current state into a form that causes a reaction in the integration center.
3. The **Signal** is how the current receptor state is communicated to the integration center. In living organisms the signal generally is either chemical or neural.
4. The **Integration Center** is where the signals from the changing states of the regulated variable are merged. The sensitivity of the integration center to the signal determines the set point. Integration centers are either individual cells or neural circuits and do not engage in conscious activity. They receive state signals from their receptors, which either activate neural circuitry or activate/deactivate chemical kinase pathways. The end result is an increase or decrease in messenger production and release.
5. The **Messenger** is how a response command is sent from the integration center to the effectors. In a living organism the effectors are always cells, because a cell or a product of a cell is responsible for everything that happens in the organism. The messenger can be either chemical or neural.
6. The **Effectors** carry out the body's response to the changes in the state of the regulated variable. At the organismic level effectors are either muscle or gland cells. At a cellular level they can be membrane channels or kinase pathways.
7. The **Effect** is the resulting change in the state of the regulated variable. If the state change reduces the stimulus causing the receptors to reduce their signal then the cause and effect chain slows down or stops. In this way the process becomes self-limiting because the effect reduces or stops the stimulus.

The **Set Point** is the level of regulated variable that activates a response (the threshold) in the negative feedback pathway. Activation of the pathway results in moving the state of the regulated variable back toward the set point (below threshold) where the negative feedback pathway activity effectively stops. Altering the sensitivity of the receptors or the chemical pathways in the integration center can change the set point (threshold).

The **Error Signal** is a measure of how strongly the regulated variable varies from the threshold level that triggers the negative feedback pathway. The greater the deviation from the threshold level the stronger the stimulation of the receptor and the more active the negative feedback pathway becomes. When the stimulus is below threshold the Error Signal is zero.

During normal circumstances negative feedback controls most body processes to maintain homeostasis. In contrast, positive feedback processes are activated when the body encounters special or unusual situations that require immediate and intense responses.

Positive Feedback Pathways

When the last effect increases the initial stimulus we call it a positive feedback loop. Positive feedback loops amplify the activity of the response and occur when a large, rapid response is required. Once triggered, Positive Feedback Pathways are self-stimulating because the effect they generate increases the stimulus. This amplifies the effect and rapidly moves the regulated variable away from its normal set point.

Positive feedback loops are usually a response to a special or unusual situation that requires rapid change to resolve, for example, childbirth, a blood vessel tear, or infection by a virus or bacteria. Once the situation is resolved (the baby is born, the tear is covered, the virus or bacteria has been eliminated) the positive stimulus generated by that need is removed, and the normal negative feedback gradually takes back control. Left unchecked, an active positive feedback loop will disrupt the stability of the system and lead to disease and ultimately the death of the organism.

Positive feedback loops often can be described using the general seven step model presented above, but require some modification. ***Instead of a regulated variable, a positive feedback process starts with an initial event***, for example the need to expel the fetus at the end of a pregnancy. The head of the fetus presses on the cervix activating stretch receptors in the cervical muscle (the receptors). These receptors send nerve impulses (signal) to the posterior pituitary (Integration Center), which releases oxytocin (messenger) in response. The oxytocin stimulates the smooth muscles of the uterus (effectors) to contract, which causes the head of the fetus to press harder on the cervix (effect). The stretch receptors in the cervical muscles increase the nerve impulses to the posterior pituitary, increasing the release of oxytocin, etc. ***The positive feedback loop stops when the special situation caused by the initial event is resolved*** (the fetus is expelled and the pressure on the stretch receptors of the cervix ceases).

Some positive feedback examples may require a little more thought to fill in the steps, but the point of the seven step model is not to be a perfect fit every time you use it. Its value is as a tool to help you consider all the steps and identify where you are missing key information or can explicitly determine why a specific step may not apply. Either way, it should help you better understand the process you are studying in terms of cause and effect rather than applying magical thinking.

Additional Considerations Regarding Homeostasis

You need to learn to think in terms of cause-effect chains and loops because much of the material will be presented in this format. As you are learning specific processes you also want to ask yourself questions about how changes at various steps would affect the process as a whole. What would happen to the process if a step were missing? What would happen to the process if a step produced a larger effect (output) than normal? This kind of thinking is important for understanding abnormal and disease states.

Homeostasis (Dynamic Kinetic Stability) is a key property of life. Even though cells are not self aware or conscious, the mindless processes of homeostasis generate the primordial characteristics of these higher-level functions. Homeostasis makes no sense without a self with a state to be maintained. The mindless activation of the negative feedback pathways that maintain homeostasis is a primitive form of the drive to survive. The homeostatic process itself follows a simple pattern that is replicated through all the levels of organization in living things: existing state – outside change – response state. The response state becomes the new existing state and the process flow of life repeats itself.

Feed Forward Pathways

Feed forward pathways differ from feedback pathways in that an activated feed forward pathway generates its effect without responding to the result. They often are activated by a change in a variable outside the feedback system that is used to anticipate a change in the the regulated variable. For example, a feed forward system causes an individual's heart rate to increase in anticipation of physical activity. The body anticipates the need for increased blood pressure before the internal pressure sensors are stimulated.

An action potential generated by a neuron can also be viewed as a feed forward pathway. If threshold is reached and the action potential is generated in the axon hillock a set of feed forward reactions transmits the potential to the axon terminals. The initiation of the original potential may be controlled by a negative feedback reflex, but once initiated, the potential travels down the axon to the terminals in a consistent and predictable cascade of reactions that results in the release of the neurotransmitter. The neurotransmitter released at the end of the axon does not directly affect the triggering of the next action potential in the neuron that released it. There is no direct feedback.

Another example is found in gene regulation where, once activated, the products of one gene become transcription factors for the next gene in the pathway. Again the key point here is that the resulting output of the pathway does not directly affect the initial stimulus. Feed forward pathways of this type are observed during cell specialization where irreversible commitments are made, such as during the development of different types of blood cells.

Feed forward systems are faster and more efficient when the triggers and appropriate responses are consistent. However, because feed forward systems are not responding directly to the state of the regulated variable they can generate inappropriate responses. For example, the social and mental challenges of modern life can trigger the same stress response as anticipating physical activity and inappropriately raise heart rate and blood pressure for prolonged periods. Ultimately this can have a negative rather than positive effect on the body.

Other Common Processes

The first is the simple process of **Input → Integration → Output**. The discussion of homeostasis above is one example of this process. However you will encounter this process in every aspect of physiology. For example, this is a seminal process in understanding how the nervous system works from the operation of a single neuron, to simple neuronal reflexes, to complex mental activity. While it may seem simple it can help you organize and order the steps in many processes.

The second is the process of **Tone**. Tone results from the equal balance between two active forces. You should have encountered this process in anatomy when you learned about muscle tone created by the opposing forces of antagonistic muscles. This balance of opposing forces not only occurs in muscles but in many other physiological processes as well. The antagonistic balance of tone allows for much finer control over a process than either on/off or graded responses from a single direction.

When homeostasis is first introduced, as in the sections above, it often is presented as an on/off process to help simplify the concepts. However, in reality most homeostatic processes are never completely off. Antagonistic processes share the same regulated variable but operate via independent negative feedback loops. Each antagonist is activated by state deviations in opposite directions. Therefore, their responses drive the regulated variable in opposite directions. Both are active thus creating the antagonistic tone. The overall direction of change in the organism is determined by small changes in the activity on one side relative to the other. You should keep the concept of tone in mind when you study the endocrine system and the autonomic system both of which maintain homeostasis to changing conditions using this mechanism.

A correlate of the balancing act created by tone is that similar results often can be obtained from two different pathways. An example is the tone created between a stimulatory process and an inhibitory process on the same regulated effector. When the antagonistic regulating forces are in balance the effector output can be increased by either increasing the stimulation or reducing the inhibition. Likewise, the the effector output can be decreased by either increasing the inhibition or decreasing the stimulus.

Because the two regulatory pathways are separate activating both at the same time to make a change can create results of a greater magnitude than predicted by summing expected results of both pathways. For example, if one pathway decreases the inhibition while the other increases the stimulus the resulting change can be life threatening. This is why Viagra and Cialis warn their potential customers not to take their product with nitrates.

Finally, several of the following chapters will cover cell membranes and several generic types of membrane proteins in detail. Keep in mind that these generic protein types are reused and modified in various ways to carry out most of the processes mentioned above. Many also are the targets of various drugs, both pharmaceutical and “street” acquired. Learning the generic protein toolkit available to cells will help you understand more complex processes presented in later chapters.

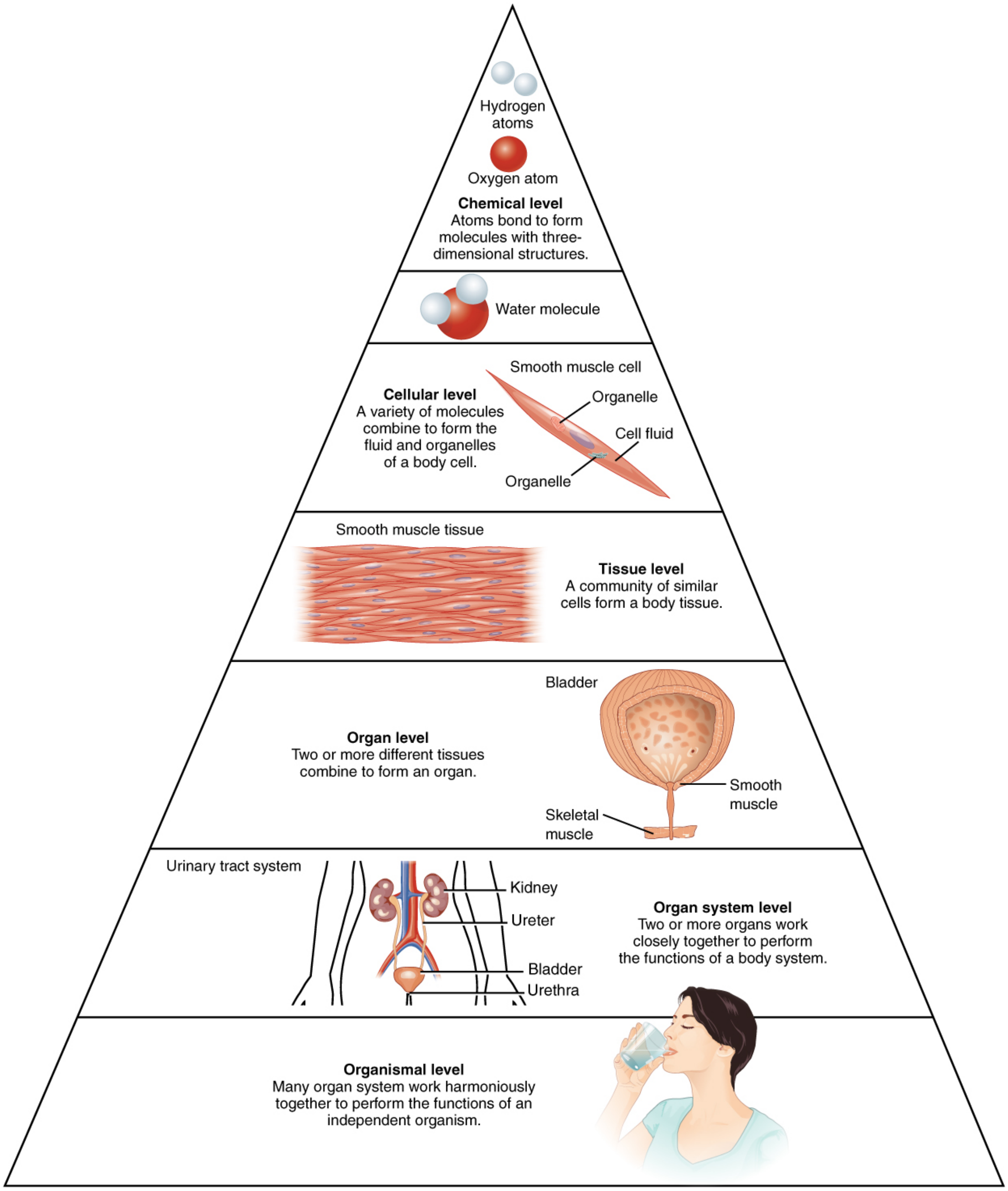
Levels of Organization

By the end of this section, you will be able to:

- Describe the structure of the human body in terms of six levels of organization
- List the eleven organ systems of the human body and identify at least one organ and one major function of each

In the study of physiology it is convenient to consider the systems of the body in terms of fundamental levels of organization that increase in complexity: subatomic particles, atoms, molecules, organelles, cells, tissues, organs, organ systems, organisms, populations, ecosystems, and biosphere ([link]).

Levels of Structural Organization of the Human Body


Hydrogen atoms
Oxygen atom
Chemical level
Atoms bond to form molecules with three-dimensional structures.
Water molecule
Smooth muscle cell
Cellular level
A variety of molecules combine to form the fluid and organelles of a body cell.
Organelle
Cell fluid
Organelle
Smooth muscle tissue
Tissue level
A community of similar cells form a body tissue.
Bladder
Organ level
Two or more different tissues combine to form an organ.
Smooth muscle
Skeletal muscle
Urinary tract system
Kidney
Ureter
Bladder
Urethra
Organ system level
Two or more organs work closely together to perform the functions of a body system.
Organismal level
Many organ system work harmoniously together to perform the functions of an independent organism.

The organization of the body often is discussed in terms of six distinct levels of increasing complexity, from the smallest chemical building blocks to a unique human organism.

Each level of organization has two characteristics you should remember.

First, each level has properties that cannot be predicted using only the rules that apply to the level below it. An easy to understand example would be to look at the Population level of organization which, if it was shown, would be immediately below the Organismal level in Figure 1 above. If you were asked, as an individual organism, "What is your birth rate?" How would you answer? You could not give a serious answer because you are only born once, so the concept of birth rate makes no sense. However, if the same question was asked about the birth rate of your class or your town, in other words a population to which you belong, a serious answer can be given. Birth rate is one of many **emergent properties** of populations. Each level of organization has emergent properties not found in the levels below it.

Second, the properties of any level of organization do not violate the rules of the levels below them. It is the rules at each level that create predictability in the entities at that level. Because of their predictable behavior they can be used as building blocks for the higher levels of organization. It is the organized relationships among these building blocks that give rise to emergent properties at the higher level. For example, the dynamic kinetic stability (DKS) of structures discussed in the previous section does not violate the laws of thermodynamics. The DKS structures exist because their constituent entities predictability follow the thermodynamic laws. Their emergent properties arise from the behavior of the organized whole. To summarize, emergent properties result from the specific order and relationships of the entities from lower levels of organization, not from making those entities operate by different rules.

The Levels of Organization

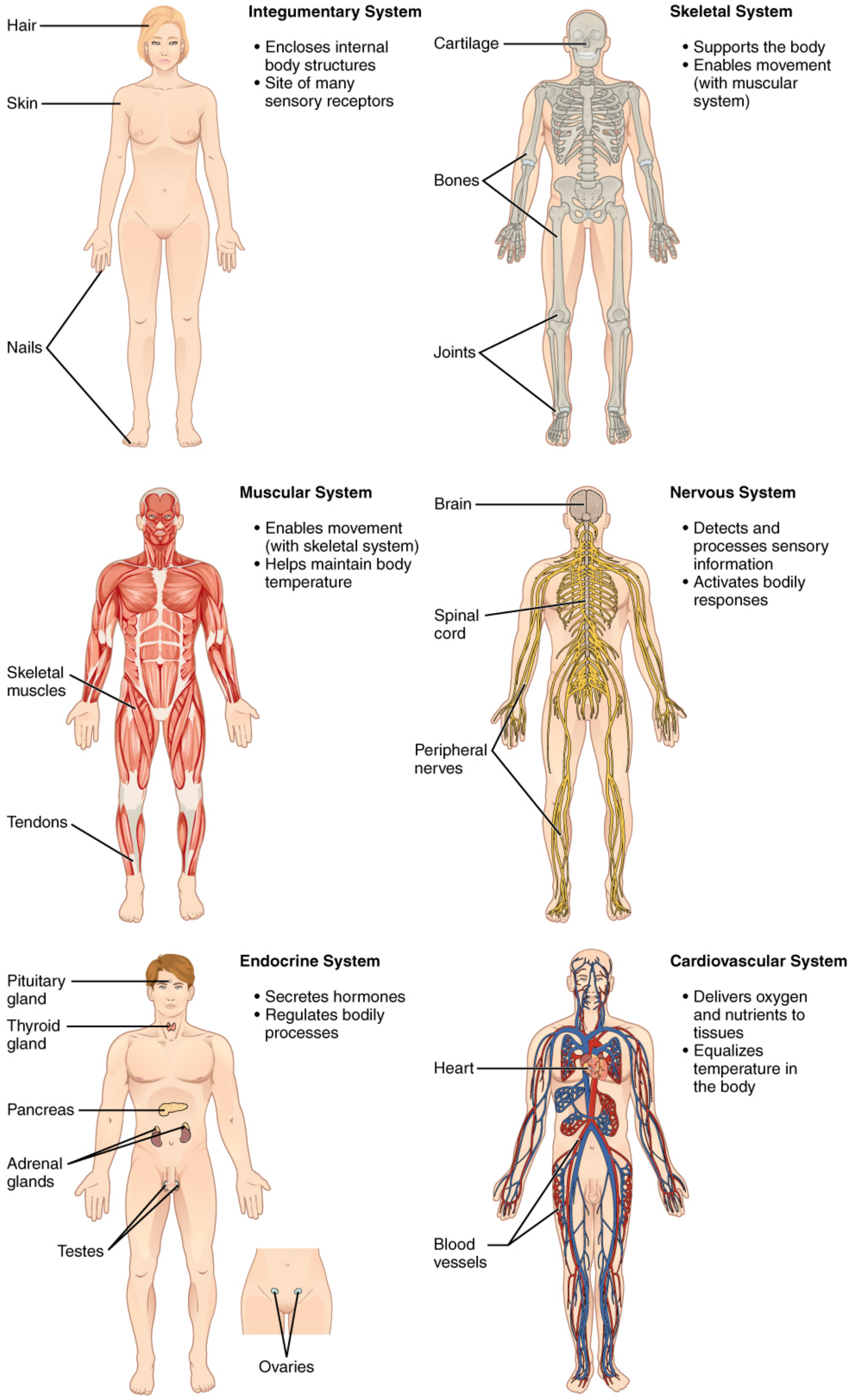
To study the chemical level of organization, scientists consider the simplest building blocks of matter: subatomic particles, atoms and molecules. All matter in the universe is composed of one or more unique pure substances called elements, familiar examples of which are hydrogen, oxygen, carbon, nitrogen, calcium, and iron. The smallest unit of any of these pure substances (elements) is an atom. Atoms are made up of subatomic particles such as the proton, electron and neutron. Two or more atoms combine to form a molecule, such as the water molecules, proteins, and sugars found in living things. Molecules are the chemical building blocks of all body structures.

A **cell** is the smallest independently functioning unit of a living organism. Even bacteria, which are extremely small, independently-living organisms, have a cellular structure. Each bacterium is a single cell. All living structures of human anatomy contain cells, and almost all functions of human physiology are performed in cells or are initiated by cells.

A human cell typically consists of flexible membranes that enclose cytoplasm, a water-based cellular fluid together with a variety of tiny functioning units called **organelles**. In humans, as in all organisms, cells perform all functions of life. A **tissue** is a group of many similar cells (though sometimes composed of a few related types) that work together to perform a specific function. An **organ** is an anatomically distinct structure of the body composed of two or more tissue types. Each organ performs one or more specific physiological functions. An **organ system** is a group of organs that work together to perform major functions or meet physiological needs of the body.

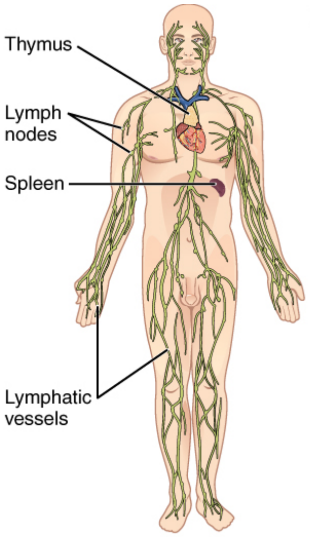
From a functional perspective there are eleven distinct organ systems in the human body ([link] and [link]). Assigning organs to organ systems can be imprecise since organs that “belong” to one system can also have functions integral to another system. In fact, most organs contribute to more than one system/function.

Organ Systems of the Human Body


Integumentary System
• Encloses internal body structures
• Site of many sensory receptors
Hair
Skin
Nails
Skeletal System
• Supports the body
• Enables movement (with muscular system)
Cartilage
Bones
Joints
Muscular System
• Enables movement (with skeletal system)
• Helps maintain body temperature
Skeletal muscles
Tendons
Nervous System
• Detects and processes sensory information
• Activates bodily responses
Brain
Spinal cord
Peripheral nerves
Endocrine System
• Secretes hormones
• Regulates bodily processes
Pituitary gland
Thyroid gland
Pancreas
Adrenal glands
Testes
Ovaries
Cardiovascular System
• Delivers oxygen and nutrients to tissues
• Equalizes temperature in the body
Heart
Blood vessels

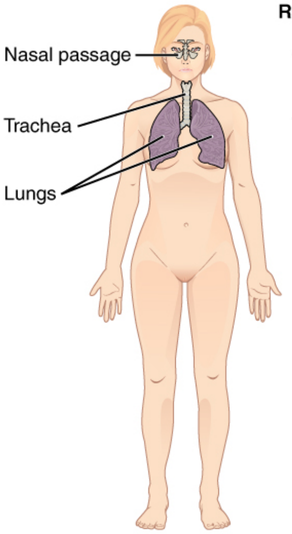
Organs that work together are grouped into organ systems.

Organ Systems of the Human Body (continued)



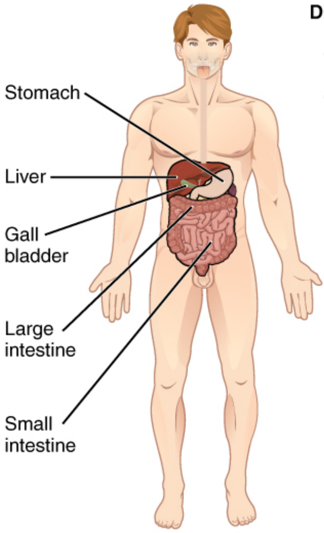
Lymphatic System

- Returns fluid to blood
- Defends against pathogens



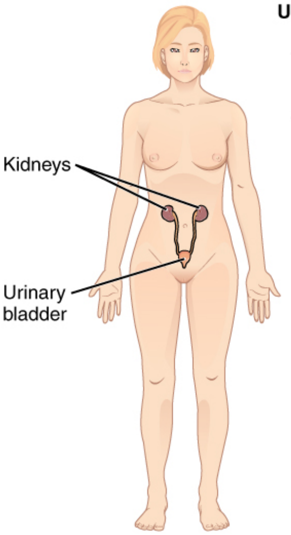
Respiratory System

- Removes carbon dioxide from the body
- Delivers oxygen to blood



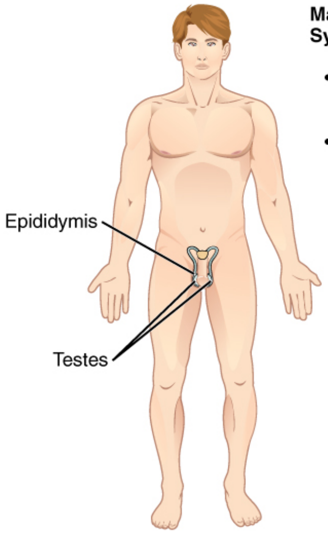
Digestive System

- Processes food for use by the body
- Removes wastes from undigested food



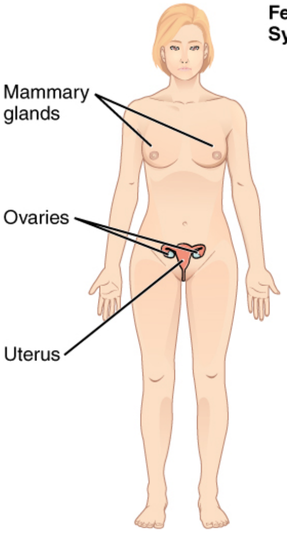
Urinary System

- Controls water balance in the body
- Removes wastes from blood and excretes them



Male Reproductive System

- Produces sex hormones and gametes
- Delivers gametes to female



Female Reproductive System

- Produces sex hormones and gametes
- Supports embryo/fetus until birth
- Produces milk for infant

Organs that work together are grouped into organ systems.

The organism level is the highest level of organization we will study in this class. An **organism** is a living being that has a cellular structure and that can independently perform all physiologic functions necessary for life. In multicellular organisms, including humans, all cells, tissues, organs, and organ systems of the body work together to maintain the life and health of the organism.

Be aware that physiology research does not stop at the organism level. Physiological principles are important at higher levels of organization. For example, physiological research is conducted at the population level in epidemiology and public health. Research at the community level is being conducted on the symbiotic role of bacteria, parasites, and viruses in maintaining the physical and mental health of individuals. And, environmental health and occupational health research rely on properties from the ecosystem level.

Chapter Review

Life processes of the human body are maintained at several levels of structural organization. These include the chemical, cellular, tissue, organ, organ system, and the organism level. Higher levels of organization are built from lower levels. Therefore, molecules combine to form cells, cells combine to form tissues, tissues combine to form organs, organs combine to form organ systems, and organ systems combine to form organisms.

Review Questions

Exercise:

Problem:

The smallest independently functioning unit of an organism is a(n) ________.

a. cell
b. molecule
c. organ
d. tissue

Solution:

A

Exercise:

Problem:

A collection of similar tissues that performs a specific function is an ________.

a. organ
b. organelle
c. organism
d. organ system

Solution:

A

Exercise:

Problem:

The body system responsible for structural support and movement is the ________.

a. cardiovascular system

b. endocrine system
c. muscular system
d. skeletal system

Solution:

D

CRITICAL THINKING QUESTIONS

Exercise:

Problem: Name the six levels of organization of the human body.

Solution:

Chemical, cellular, tissue, organ, organ system, organism.

Exercise:

Problem:

The female ovaries and the male testes are a part of which body system? Can these organs be members of more than one organ system? Why or why not?

Solution:

The female ovaries and the male testes are parts of the reproductive system. But they also secrete hormones, as does the endocrine system, therefore ovaries and testes function within both the endocrine and reproductive systems.

Glossary

cell

smallest independently functioning unit of all organisms; in animals, a cell contains cytoplasm, composed of fluid and organelles

organ

functionally distinct structure composed of two or more types of tissues

organ system

group of organs that work together to carry out a particular function

organism

living being that has a cellular structure and that can independently perform all physiologic functions necessary for life

tissue

group of similar or closely related cells that act together to perform a specific function

Functions of Human Life
By the end of this section, you will be able to:

- Explain the importance of organization to the function of the human organism
- Distinguish between metabolism, anabolism, and catabolism
- Provide at least two examples of human responsiveness and human movement
- Compare and contrast growth, differentiation, and reproduction

The different organ systems each have different functions and therefore unique roles to perform in physiology. These many functions can be summarized in terms of a few that we might consider definitive of human life: organization, metabolism, responsiveness, movement, development, and reproduction.

Organization

A human body consists of trillions of cells organized in a way that maintains distinct internal compartments. These compartments keep body cells separated from external environmental threats and keep the cells moist and nourished. They also separate internal body fluids from the countless microorganisms that grow on body surfaces, including the lining of certain tracts, or passageways. The intestinal tract, for example, is home to even more bacteria cells than the total of all human cells in the body, yet these bacteria are outside the body and cannot be allowed to circulate freely inside the body.

Cells, for example, have a cell membrane (also referred to as the plasma membrane) that keeps the intracellular environment—the fluids and organelles—separate from the extracellular environment. Blood vessels keep blood inside a closed circulatory system, and nerves and muscles are wrapped in connective tissue sheaths that separate them from surrounding structures. In the chest and abdomen, a variety of internal membranes keep major organs such as the lungs, heart, and kidneys separate from others.

The body’s largest organ system is the integumentary system, which includes the skin and its associated structures, such as hair and nails. The surface tissue of skin is a barrier that protects internal structures and fluids from potentially harmful microorganisms and other toxins.

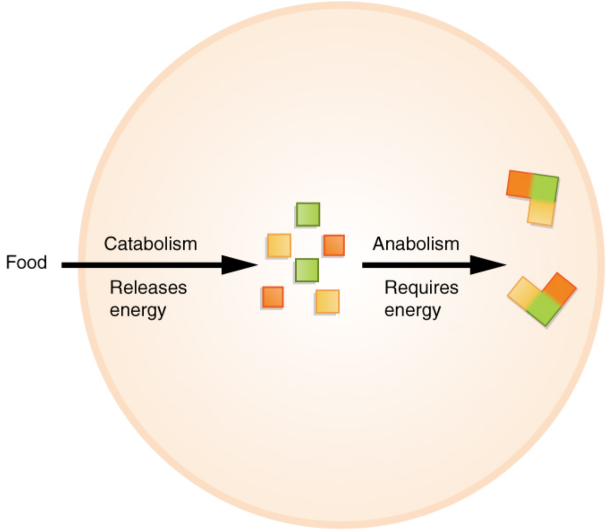
Metabolism

The first law of thermodynamics holds that energy can neither be created nor destroyed—it can only change form. Your basic function as an organism is to consume (ingest) energy and molecules in the foods you eat, convert some of it into fuel for movement, sustain your body functions, and build and maintain your body structures. There are two types of reactions that accomplish this: **anabolism** and **catabolism**.

- **Anabolism** is the process whereby smaller, simpler molecules are combined into larger, more complex substances. Your body can assemble, by utilizing energy, the complex chemicals it needs by combining small molecules derived from the foods you eat
- **Catabolism** is the process by which larger more complex substances are broken down into smaller simpler molecules. Catabolism releases energy. The complex molecules found in foods are broken down so the body can use their parts to assemble the structures and substances needed for life.

Taken together, these two processes are called metabolism. **Metabolism** is the sum of all anabolic and catabolic reactions that take place in the body ([link]). Both anabolism and catabolism occur simultaneously and continuously to keep you alive.

Metabolism



Anabolic reactions are building reactions, and they consume energy. Catabolic reactions break materials down and release energy. Metabolism includes both anabolic and catabolic reactions.

Every cell in your body makes use of a chemical compound, **adenosine triphosphate (ATP)**, to store and release energy. The cell stores energy in the synthesis (anabolism) of ATP, then moves the ATP molecules to the location where energy is needed to fuel cellular activities. Then the ATP is broken down (catabolism) and a controlled amount of energy is released, which is used by the cell to perform a particular job.

Note:

View this animation to learn more about metabolic processes. Which organs of the body likely carry out anabolic processes? What about catabolic processes?

Responsiveness

Responsiveness is the ability of an organism to adjust to changes in its internal and external environments. An example of responsiveness to external stimuli could include moving toward sources of food and water and away from perceived dangers. Changes in an organism's internal environment, such as increased body temperature, can cause the responses of sweating and the dilation of blood vessels in the skin in order to decrease body temperature, as shown by the runners in [link].

Movement

Human movement includes not only actions at the joints of the body, but also the motion of individual organs and even individual cells. As you read these words, red and white blood cells are moving throughout your body, muscle cells are contracting and relaxing to maintain your posture and to focus your vision, and glands are secreting chemicals to regulate body functions. Your body is coordinating the action of entire muscle groups to enable you to move air into and out of your lungs, to push blood throughout your body, and to propel the food you have eaten through your digestive tract. Consciously, of course, you contract your skeletal muscles to move the bones of your skeleton to get from one place to another (as the runners are doing in [link]), and to carry out all of the activities of your daily life.

Marathon Runners

Runners demonstrate two characteristics of living humans—responsiveness and movement. Anatomic structures and physiological processes allow runners to coordinate the action of muscle groups and sweat in response to rising internal body temperature. (credit: Phil Roeder/flickr)

Development, growth and reproduction

Development is all of the changes the body goes through in life. Development includes the process of **differentiation**, in which unspecialized cells become specialized in structure and function to perform certain tasks in the body. Development also includes the processes of growth and repair, both of which involve cell differentiation.

Growth is the increase in body size. Humans, like all multicellular organisms, grow by increasing the number of existing cells, increasing the amount of non-cellular material around cells (such as mineral deposits in bone), and, within very narrow limits, increasing the size of existing cells.

Reproduction is the formation of a new organism from parent organisms. In humans, reproduction is carried out by the male and female reproductive systems. Because death will come to all complex organisms, without reproduction, the line of organisms would end.

Chapter Review

Most processes that occur in the human body are not consciously controlled. They occur continuously to build, maintain, and sustain life. These processes include: organization, in terms of the maintenance of essential body boundaries; metabolism, including energy transfer via anabolic and catabolic reactions; responsiveness; movement; and growth, differentiation, reproduction, and renewal.

Interactive Link Questions

Exercise:

Problem:

View this animation to learn more about metabolic processes. What kind of catabolism occurs in the heart?

Solution:

Fatty acid catabolism.

Review Questions

Exercise:

Problem: Metabolism can be defined as the _________.

a. adjustment by an organism to external or internal changes
b. process whereby all unspecialized cells become specialized to perform distinct functions

c. process whereby new cells are formed to replace worn-out cells
d. sum of all chemical reactions in an organism

Solution:

D

Exercise:

Problem:

Adenosine triphosphate (ATP) is an important molecule because it ________.

a. is the result of catabolism
b. release energy in uncontrolled bursts
c. stores energy for use by body cells
d. All of the above

Solution:

C

Exercise:

Problem:

Cancer cells can be characterized as “generic” cells that perform no specialized body function. Thus cancer cells lack ________.

a. differentiation
b. reproduction
c. responsiveness
d. both reproduction and responsiveness

Solution:

A

CRITICAL THINKING QUESTIONS

Exercise:

Problem:

Explain why the smell of smoke when you are sitting at a campfire does not trigger alarm, but the smell of smoke in your residence hall does.

Solution:

When you are sitting at a campfire, your sense of smell adapts to the smell of smoke. Only if that smell were to suddenly and dramatically intensify would you be likely to notice and respond. In contrast, the smell of even a trace of smoke would be new and highly unusual in your residence hall, and would be perceived as danger.

Exercise:

Problem:

Identify three different ways that growth can occur in the human body.

Solution:

Growth can occur by increasing the number of existing cells, increasing the size of existing cells, or increasing the amount of non-cellular material around cells.

Glossary

anabolism	assembly of more complex molecules from simpler molecules
catabolism	breaking down of more complex molecules into simpler molecules

development
changes an organism goes through during its life

differentiation
process by which unspecialized cells become specialized in structure and function

growth
process of increasing in size

metabolism
sum of all of the body’s chemical reactions

renewal
process by which worn-out cells are replaced

reproduction
process by which new organisms are generated

responsiveness
ability of an organisms or a system to adjust to changes in conditions

Requirements for Human Life

By the end of this section, you will be able to:

- Discuss the role of oxygen and nutrients in maintaining human survival
- Explain why extreme heat and extreme cold threaten human survival
- Explain how the pressure exerted by gases and fluids influences human survival

Humans have been adapting to life on Earth for at least the past 200,000 years. Earth and its atmosphere have provided us with air to breathe, water to drink, and food to eat, but these are not the only requirements for survival. Although you may rarely think about it, you also cannot live outside of a certain range of temperature and pressure that the surface of our planet and its atmosphere provides. The next sections explore these four requirements of life.

Oxygen

Atmospheric air is only about 20 percent oxygen, but that oxygen is a key component of the chemical reactions that keep the body alive, including the reactions that produce ATP. Brain cells are especially sensitive to lack of oxygen because of their requirement for a high-and-steady production of ATP. Brain damage is likely within five minutes without oxygen, and death is likely within ten minutes.

Nutrients

A **nutrient** is a substance in foods and beverages that is essential to human survival. The three basic classes of nutrients are water, the energy-yielding and body-building nutrients, and the micronutrients (vitamins and minerals).

The most critical nutrient is water. Depending on the environmental temperature and our state of health, we may be able to survive for only a few days without water. The body’s functional chemicals are dissolved and transported in water, and the chemical reactions of life take place in water.

Moreover, water is the largest component of cells, blood, and the fluid between cells, and water makes up about 70 percent of an adult's body mass. Water also helps regulate our internal temperature and cushions, protects, and lubricates joints and many other body structures.

The energy-yielding nutrients are primarily carbohydrates and lipids, while proteins mainly supply the amino acids that are the building blocks of the body itself. You ingest these in plant and animal foods and beverages, and the digestive system breaks them down into molecules small enough to be absorbed. The breakdown products of carbohydrates and lipids can then be used in the metabolic processes that convert them to ATP. Although you might feel as if you are starving after missing a single meal, you can survive without consuming the energy-yielding nutrients for at least several weeks.

Water and the energy-yielding nutrients are also referred to as macronutrients because the body needs them in large amounts. In contrast, micronutrients are vitamins and minerals. These elements and compounds participate in many essential chemical reactions and processes, such as nerve impulses, and some, such as calcium, also contribute to the body's structure. Your body can store some of the micronutrients in its tissues, and draw on those reserves if you fail to consume them in your diet for a few days or weeks. Some others micronutrients, such as vitamin C and most of the B vitamins, are water-soluble and cannot be stored, so you need to consume them every day or two.

Narrow Range of Temperature

You have probably seen news stories about athletes who died of heat stroke, or hikers who died of exposure to cold. Such deaths occur because the chemical reactions upon which the body depends can only take place within a narrow range of body temperature, from just below to just above 37°C (98.6°F). When body temperature rises well above or drops well below normal, certain proteins (enzymes) that facilitate chemical reactions lose their normal structure and their ability to function and the chemical reactions of metabolism cannot proceed.

That said, the body can respond effectively to short-term exposure to heat ([link]) or cold. One of the body’s responses to heat is, of course, sweating. As sweat evaporates from skin, it removes some thermal energy from the body, cooling it. Adequate water (from the extracellular fluid in the body) is necessary to produce sweat, so adequate fluid intake is essential to balance that loss during the sweat response. Not surprisingly, the sweat response is much less effective in a humid environment because the air is already saturated with water. Thus, the sweat on the skin’s surface is not able to evaporate, and internal body temperature can get dangerously high.

Extreme Heat

Humans adapt to some degree to repeated exposure to high temperatures. (credit: McKay Savage/flickr)

The body can also respond effectively to short-term exposure to cold. One response to cold is shivering, which is random muscle movement that generates heat. Another response is increased breakdown of stored energy to generate heat. When that energy reserve is depleted, however, and the core temperature begins to drop significantly, red blood cells will lose their ability to give up oxygen, denying the brain of this critical component of ATP production. This lack of oxygen can cause confusion, lethargy, and eventually loss of consciousness and death. The body responds to cold by reducing blood circulation to the extremities, the hands and feet, in order to prevent blood from cooling there and so that the body’s core can stay warm. Even when core body temperature remains stable, however, tissues exposed to severe cold, especially the fingers and toes, can develop frostbite when blood flow to the extremities has been much reduced. This form of tissue damage can be permanent and lead to gangrene, requiring amputation of the affected region.

Note:

Everyday Connection

Controlled Hypothermia

As you have learned, the body continuously engages in coordinated physiological processes to maintain a stable temperature. In some cases, however, overriding this system can be useful, or even life-saving. Hypothermia is the clinical term for an abnormally low body temperature (hypo- = “below” or “under”). Controlled hypothermia is clinically induced hypothermia performed in order to reduce the metabolic rate of an organ or of a person’s entire body.

Controlled hypothermia often is used, for example, during open-heart surgery because it decreases the metabolic needs of the brain, heart, and other organs, reducing the risk of damage to them. When controlled hypothermia is used clinically, the patient is given medication to prevent shivering. The body is then cooled to 25–32°C (79–89°F). The heart is stopped and an external heart-lung pump maintains circulation to the patient’s body. The heart is cooled further and is maintained at a temperature below 15°C (60°F) for the duration of the surgery. This very cold temperature helps the heart muscle to tolerate its lack of blood supply during the surgery.

Some emergency department physicians use controlled hypothermia to reduce damage to the heart in patients who have suffered a cardiac arrest. In the emergency department, the physician induces coma and lowers the patient’s body temperature to approximately 91 degrees. This condition, which is maintained for 24 hours, slows the patient’s metabolic rate. Because the patient’s organs require less blood to function, the heart’s workload is reduced.

Narrow Range of Atmospheric Pressure

Pressure is a force exerted by a substance that is in contact with another substance. Atmospheric pressure is pressure exerted by the mixture of gases (primarily nitrogen and oxygen) in the Earth's atmosphere. Although you may not perceive it, atmospheric pressure is constantly pressing down on your body. This pressure keeps gases within your body, such as the gaseous nitrogen in body fluids, dissolved. If you were suddenly ejected from a space ship above Earth's atmosphere, you would go from a situation of normal pressure to one of very low pressure. The pressure of the nitrogen gas in your blood would be much higher than the pressure of nitrogen in the space surrounding your body. As a result, the nitrogen gas in your blood would expand, forming bubbles that could block blood vessels and even cause cells to break apart.

Atmospheric pressure does more than just keep blood gases dissolved. Your ability to breathe—that is, to take in oxygen and release carbon dioxide—also depends upon a precise atmospheric pressure. Altitude sickness occurs in part because the atmosphere at high altitudes exerts less pressure, reducing the exchange of these gases, and causing shortness of breath, confusion, headache, lethargy, and nausea. Mountain climbers carry oxygen to reduce the effects of both low oxygen levels and low barometric pressure at higher altitudes ([link]).

Harsh Conditions

Climbers on Mount Everest must

accommodate extreme cold, low oxygen levels, and low barometric pressure in an environment hostile to human life. (credit: Melanie Ko/flickr)

Note:

Homeostatic Imbalances

Decompression Sickness

Decompression sickness (DCS) is a condition in which gases dissolved in the blood or in other body tissues are no longer dissolved following a reduction in pressure on the body. This condition affects underwater divers who surface from a deep dive too quickly, and it can affect pilots flying at high altitudes in planes with unpressurized cabins. Divers often call this condition “the bends,” a reference to joint pain that is a symptom of DCS.

In all cases, DCS is brought about by a reduction in barometric pressure. At high altitude, barometric pressure is much less than on Earth’s surface because pressure is produced by the weight of the column of air above the body pressing down on the body. The very great pressures on divers in deep water are likewise from the weight of a column of water pressing down on the body. For divers, DCS occurs at normal barometric pressure (at sea level), but it is brought on by the relatively rapid decrease of pressure as divers rise from the high pressure conditions of deep water to the now low, by comparison, pressure at sea level. Not surprisingly, diving in deep mountain lakes, where barometric pressure at the surface of the lake is less than that at sea level is more likely to result in DCS than diving in water at sea level.

In DCS, gases dissolved in the blood (primarily nitrogen) come rapidly out of solution, forming bubbles in the blood and in other body tissues. This occurs because when pressure of a gas over a liquid is decreased, the amount of gas that can remain dissolved in the liquid also is decreased. It is air pressure that keeps your normal blood gases dissolved in the blood. When pressure is reduced, less gas remains dissolved. You have seen this in effect when you open a carbonated drink. Removing the seal of the bottle reduces the pressure of the gas over the liquid. This in turn causes bubbles as dissolved gases (in this case, carbon dioxide) come out of solution in the liquid.

The most common symptoms of DCS are pain in the joints, with headache and disturbances of vision occurring in 10 percent to 15 percent of cases. Left untreated, very severe DCS can result in death. Immediate treatment is with pure oxygen. The affected person is then moved into a hyperbaric chamber. A hyperbaric chamber is a reinforced, closed chamber that is pressurized to greater than atmospheric pressure. It treats DCS by repressurizing the body so that pressure can then be removed much more gradually. Because the hyperbaric chamber introduces oxygen to the body at high pressure, it increases the concentration of oxygen in the blood. This has the effect of replacing some of the nitrogen in the blood with oxygen, which is easier to tolerate out of solution.

The dynamic pressure of body fluids is also important to human survival. For example, blood pressure, which is the pressure exerted by blood as it flows within blood vessels, must be great enough to enable blood to reach all body tissues, and yet low enough to ensure that the delicate blood vessels can withstand the friction and force of the pulsating flow of pressurized blood.

Chapter Review

Humans cannot survive for more than a few minutes without oxygen, for more than several days without water, and for more than several weeks without carbohydrates, lipids, proteins, vitamins, and minerals. Although the body can respond to high temperatures by sweating and to low temperatures by shivering and increased fuel consumption, long-term exposure to extreme heat and cold is not compatible with survival. The body requires a precise atmospheric pressure to maintain its gases in solution and to facilitate respiration—the intake of oxygen and the release of carbon dioxide. Humans also require blood pressure high enough to ensure that blood reaches all body tissues but low enough to avoid damage to blood vessels.

Review Questions

Exercise:

Problem:

Humans have the most urgent need for a continuous supply of ________.

a. food
b. nitrogen
c. oxygen
d. water

Solution:

C

Exercise:

Problem: Which of the following statements about nutrients is true?

a. All classes of nutrients are essential to human survival.
b. Because the body cannot store any micronutrients, they need to be consumed nearly every day.
c. Carbohydrates, lipids, and proteins are micronutrients.
d. Macronutrients are vitamins and minerals.

Solution:

A

Exercise:

Problem:

C.J. is stuck in her car during a bitterly cold blizzard. Her body responds to the cold by ________.

a. increasing the blood to her hands and feet
b. becoming lethargic to conserve heat
c. breaking down stored energy
d. significantly increasing blood oxygen levels

Solution:

C

CRITICAL THINKING QUESTIONS

Exercise:

Problem:

When you open a bottle of sparkling water, the carbon dioxide gas in the bottle form bubbles. If the bottle is left open, the water will eventually “go flat.” Explain these phenomena in terms of atmospheric pressure.

Solution:

In a sealed bottle of sparkling water, carbon dioxide gas is kept dissolved in the water under a very high pressure. When you open the bottle, the pressure of the gas above the liquid changes from artificially high to normal atmospheric pressure. The dissolved carbon dioxide gas expands, and rises in bubbles to the surface. When a bottle of sparkling water is left open, it eventually goes flat because its gases continue to move out of solution until the pressure in the water is approximately equal to atmospheric pressure.

Exercise:

Problem:

On his midsummer trek through the desert, Josh ran out of water. Why is this particularly dangerous?

Solution:

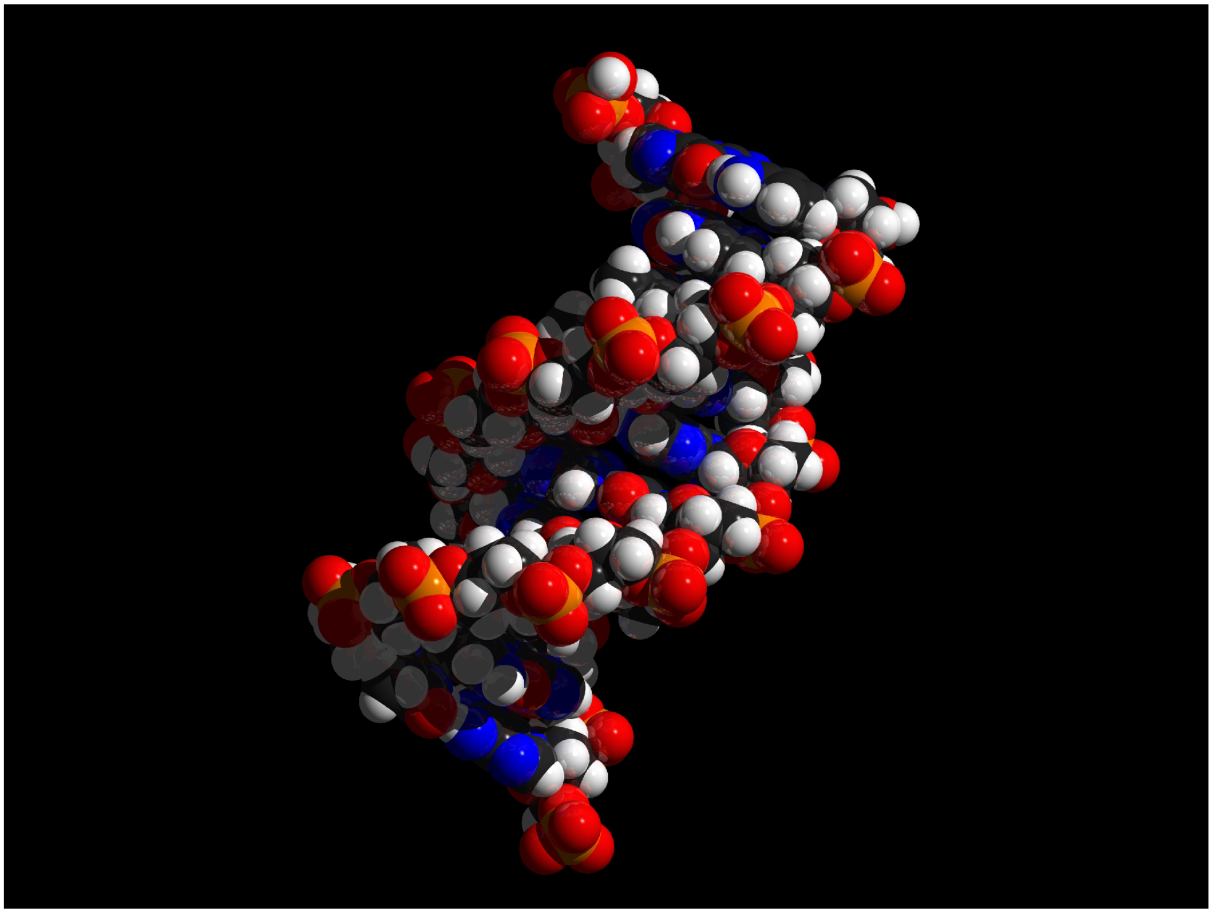
The primary way that the body responds to high environmental heat is by sweating; however, sweating requires water, which comes from body fluids, including blood plasma. If Josh becomes dehydrated, he will be unable to sweat adequately to cool his body, and he will be at risk for heat stroke as his blood pressure drops too much from the loss of water from the blood plasma.

Glossary

nutrient	chemical obtained from foods and beverages that is critical to human survival
pressure	force exerted by a substance in contact with another substance

Introduction
class="introduction"
Human DNA

Human DNA is described as a double helix that resembles a molecular spiral staircase. In humans the DNA is organized into 46 chromosomes.



Note:

Chapter Objectives

After studying this chapter, you will be able to:

- Describe the fundamental composition of matter
- Identify the three subatomic particles
- Identify the four most abundant elements in the body
- Explain the relationship between an atom’s number of electrons and its relative stability
- Distinguish between ionic bonds, covalent bonds, and hydrogen bonds
- Explain how energy is invested, stored, and released via chemical reactions, particularly those reactions that are critical to life
- Explain the importance of the inorganic compounds that contribute to life, such as water, salts, acids, and bases

- Compare and contrast the four important classes of organic (carbon-based) compounds—proteins, carbohydrates, lipids and nucleic acids—according to their composition and functional importance to human life

The smallest, most fundamental material components of the human body are basic chemical elements. In fact, chemicals called nucleotide bases are the foundation of the genetic code with the instructions on how to build and maintain the human body from conception through old age. There are about three billion of these base pairs in human DNA.

Human chemistry includes organic molecules (carbon-based) and biochemicals (those produced by the body). Human chemistry also includes elements. In fact, life cannot exist without many of the elements that are part of the earth. All of the elements that contribute to chemical reactions, to the transformation of energy, and to electrical activity and muscle contraction—elements that include phosphorus, carbon, sodium, and calcium, to name a few—originated in stars.

These elements, in turn, can form both the inorganic and organic chemical compounds important to life, including, for example, water, glucose, and proteins. This chapter begins by examining elements and how the structures of atoms, the basic units of matter, determine the characteristics of elements by the number of protons, neutrons, and electrons in the atoms. The chapter then builds the framework of life from there.

Fundametals of Matter: Elements and Their Relationships

By the end of this section, you will be able to:

- Define matter and elements
- Describe the interrelationship between protons, neutrons, and electrons
- Compare the ways in which electrons can be donated or shared between atoms
- Explain the ways in which naturally occurring elements combine to create molecules, cells, tissues, organ systems, and organisms

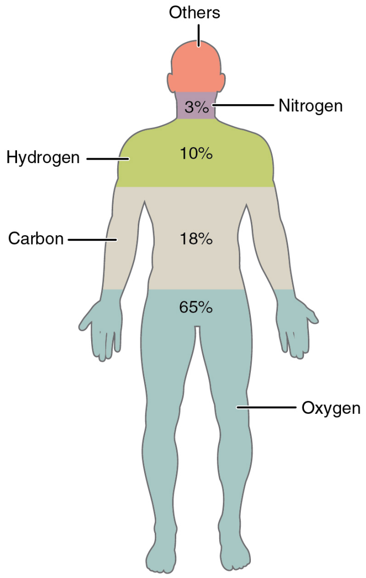
At its most fundamental level, life is made up of matter. **Matter** is any substance that occupies space and has mass. **Elements** are unique forms of matter with specific chemical and physical properties that cannot be broken down into smaller substances by ordinary chemical reactions. There are 118 elements, but only 92 occur naturally. The remaining elements are synthesized in laboratories and are unstable.

Each element is designated by its chemical symbol, which is a single capital letter or, when the first letter is already “taken” by another element, a combination of two letters. Some elements follow the English term for the element, such as C for carbon and Ca for calcium. Other elements’ chemical symbols derive from their Latin names; for example, the symbol for sodium is Na, referring to *natrium*, the Latin word for sodium.

While your body can assemble many of the chemical compounds needed for life from their constituent elements, it cannot make elements. They must come from the environment. The four elements common to all living organisms are oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). In the non-living world, elements are found in different proportions, and some elements common to living organisms are relatively rare on the earth as a whole, as shown in [link]. For example, the atmosphere is rich in nitrogen and oxygen but contains little carbon and hydrogen, while the earth’s crust, although it contains oxygen and a small amount of hydrogen, has little nitrogen and carbon. In spite of their differences in abundance, all elements and the chemical reactions between them obey the same chemical and physical laws regardless of whether they are a part of the living or non-living world.

Approximate Percentage of Elements in Living Organisms (Humans) Compared to the Non-living World			
Element	**Life (Humans)**	**Atmosphere**	**Earth’s Crust**
Oxygen (O)	65%	21%	46%
Carbon (C)	18%	trace	trace
Hydrogen (H)	10%	trace	0.1%
Nitrogen (N)	3%	78%	trace

In addition to the four most common elements found in living organisms additional elements also are required to maintain life. A familiar example of an element that you require is calcium (Ca^{++}). Calcium is essential to the human body; it is absorbed and used for a number of processes, including strengthening bones, contracting muscles, releasing neurotransmitter, and many other cellular activities. Other elements important to life that you might be familiar with are sodium, potassium, and iron. The elements in the human body are shown in [link], beginning with the most abundant: oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). Each element’s name can be replaced by a one- or two-letter symbol; you will become familiar with some of these during this course. All the elements in your body are derived from the foods you eat and the air you breathe.



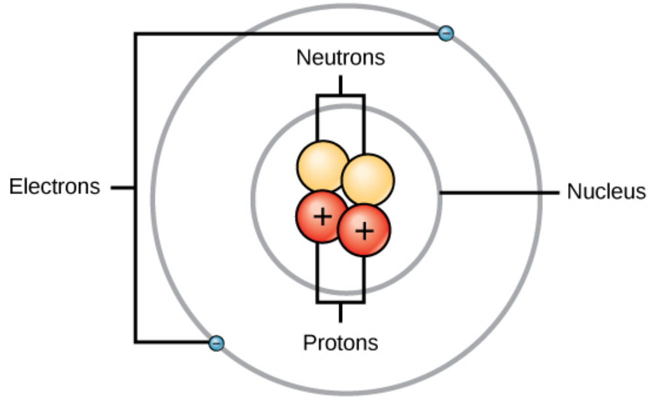
Element	Symbol	Percentage in Body
Oxygen	O	65.0
Carbon	C	18.5
Hydrogen	H	9.5
Nitrogen	N	3.2
Calcium	Ca	1.5
Phosphorus	P	1.0
Potassium	K	0.4
Sulfur	S	0.3
Sodium	Na	0.2
Chlorine	Cl	0.2
Magnesium	Mg	0.1
Trace elements include boron (B), chromium (Cr), cobalt (Co), copper (Cu), fluorine (F), iodine (I), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se), silicon (Si), tin (Sn), vanadium (V), and zinc (Zn).		less than 1.0

The main elements that compose the human body are shown from most abundant to least abundant.

The Structure of the Atom

To understand how elements come together, we must first discuss the smallest component or building block of an element, the atom. An **atom** is the smallest unit of matter that retains all of the chemical properties of an element. For example, one gold atom has all of the properties of gold in that it is a solid metal at room temperature. A gold coin is simply a very large number of gold atoms molded into the shape of a coin and containing small amounts of other elements known as impurities. Gold atoms cannot be broken down into anything smaller while still retaining the properties of gold.

An atom is composed of two regions: the **nucleus**, which is in the center of the atom and contains protons and neutrons, and the outermost region of the atom which holds its electrons in orbit around the nucleus, as illustrated in [link]. Atoms contain protons, electrons, and neutrons, among other subatomic particles. The only exception is hydrogen (H), which is made of one proton and one electron with no neutrons.



Elements, such as helium, depicted here, are made up of atoms. Atoms are made up of protons and neutrons located within the nucleus, with electrons in orbitals surrounding the nucleus.

Protons and neutrons have approximately the same mass, about 1.67 × 10^{-24} grams. Scientists arbitrarily define this amount of mass as one atomic mass unit (amu) or one Dalton, as shown in [link]. Although similar in mass, protons and neutrons differ in their electric charge. A **proton** is positively charged whereas a **neutron** is uncharged. Therefore, the number of neutrons in an atom contributes significantly to its mass, but not to its charge. **Electrons** are much smaller in mass than protons, weighing only 9.11 × 10^{-28} grams, or about 1/1800 of an atomic mass unit. Hence, they do not contribute much to an element's overall atomic mass. Therefore, when considering atomic mass, it is customary to ignore the mass of any electrons and calculate the atom's mass based on the number of protons and neutrons alone. Although not significant contributors to mass, electrons do contribute greatly to the atom's charge, as each electron has a negative charge equal to the positive charge of a proton. In uncharged, neutral atoms, the number of electrons orbiting the nucleus is equal to the number of protons inside the nucleus. In these atoms, the positive and negative charges cancel each other out, leading to an atom with no net charge.

Accounting for the sizes of protons, neutrons, and electrons, most of the volume of an atom—greater than 99 percent—is, in fact, empty space. With all this empty space, one might ask why so-called solid objects do not just pass through one another. The reason they do not is that the electrons that surround all atoms are negatively charged and negative charges repel each other.

Protons, Neutrons, and Electrons

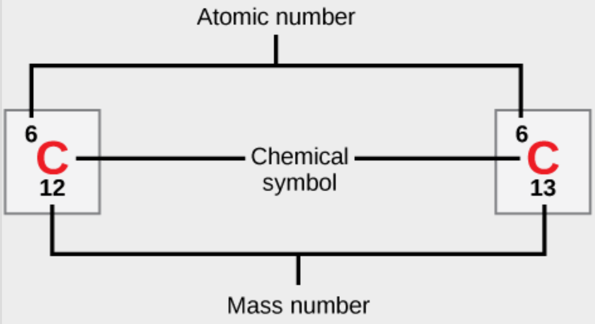
	Charge	Mass (amu)	Location
Proton	+1	1	nucleus
Neutron	0	1	nucleus
Electron	–1	0	orbitals

Atomic Number and Mass

Atoms of each element contain a characteristic number of protons and electrons. The number of protons determines an element's **atomic number** and is used to distinguish one element from another. The number of neutrons is variable, resulting in isotopes, which are different forms of the same atom that vary only in the number of neutrons they possess. Together, the number of protons and the number of neutrons determine an element's **mass number**, as illustrated in [link]. Note that the small contribution of mass from electrons is disregarded in calculating the mass number. This approximation of mass can be used to easily calculate how many neutrons an element has by simply subtracting the number of protons from the mass number. Since an element's isotopes will have slightly different mass numbers, scientists also determine the **atomic mass**, which is the calculated mean of the mass number for its naturally occurring isotopes. Often, the resulting number contains a fraction. For example, the atomic mass of chlorine (Cl) is 35.45 because chlorine is composed of several isotopes, some (the majority) with atomic mass 35 (17 protons and 18 neutrons) and some with atomic mass 37 (17 protons and 20 neutrons).

Note:

Art Connection



Carbon has an atomic number of six, and two stable isotopes with mass numbers of twelve and thirteen, respectively. Its atomic mass is 12.11.

How many neutrons do carbon-12 and carbon-13 have, respectively?

Isotopes

Isotopes are different forms of an element that have the same number of protons but a different number of neutrons. Some elements—such as carbon, potassium, and uranium—have naturally occurring isotopes. Carbon-12 contains six protons, six neutrons, and six electrons; therefore, it has a mass number of 12 (six protons and six neutrons). Carbon-14 contains six protons, eight neutrons, and six electrons; its atomic mass is 14 (six protons and eight neutrons). These two alternate forms of carbon are isotopes. Some isotopes may emit neutrons, protons, and electrons, and attain a more stable atomic configuration (lower level of potential energy); these are radioactive isotopes, or **radioisotopes**. Radioactive decay (carbon-14 losing neutrons to eventually become carbon-12) describes the energy loss that occurs when an unstable atom's nucleus releases radiation.

Note:

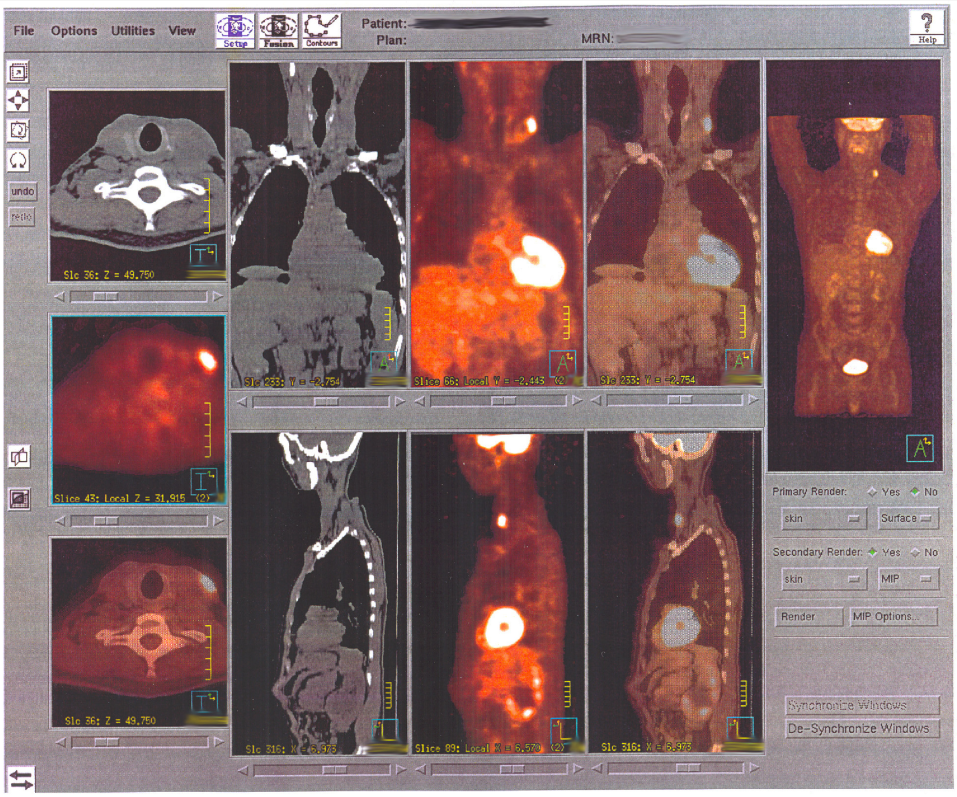
Career Connection

Interventional Radiologist

The controlled use of radioisotopes has advanced medical diagnosis and treatment of disease. Interventional radiologists are physicians who treat disease by using minimally invasive techniques involving radiation. Many conditions that could once only be treated with a lengthy and traumatic operation can now be treated non-surgically, reducing the cost, pain, length of hospital stay, and recovery time for patients. For example, in the past, the only options for a patient with one or more tumors in the liver were surgery and chemotherapy (the administration of drugs to treat cancer). Some liver tumors, however, are difficult to access surgically, and others could require the surgeon to remove too much of the liver. Moreover, chemotherapy is highly toxic to the liver, and certain tumors do not respond well to it anyway. In some such cases, an interventional radiologist can treat the tumors by disrupting their blood supply, which they need if they are to continue to grow. In this procedure, called radioembolization, the radiologist accesses the liver with a fine needle, threaded through one of the patient's blood vessels. The radiologist then inserts tiny radioactive "seeds" into the blood vessels that supply the tumors. In the days and weeks following the procedure, the radiation emitted from the seeds destroys the vessels and directly kills the tumor cells in the vicinity of the treatment.

Radioisotopes emit subatomic particles that can be detected and tracked by imaging technologies. One of the most advanced uses of radioisotopes in medicine is the positron emission tomography (PET) scanner, which detects the activity in the body of a very small injection of radioactive glucose, the simple sugar that cells use for energy. The PET camera reveals to the medical team which of the patient’s tissues are taking up the most glucose. Thus, the most metabolically active tissues show up as bright “hot spots” on the images ([link]). PET can reveal some cancerous masses because cancer cells consume glucose at a high rate to fuel their rapid reproduction.

PET Scan



Patient: [redacted]
Plan:
MRN: [redacted]

Primary Render: ○ Yes ● No
skin
Surface
Secondary Render: ● Yes ○ No
skin
MIP

PET highlights areas in the body where there is relatively high glucose use, which is characteristic of cancerous tissue. This PET scan shows sites of the spread of a large primary tumor to other sites.

Note:

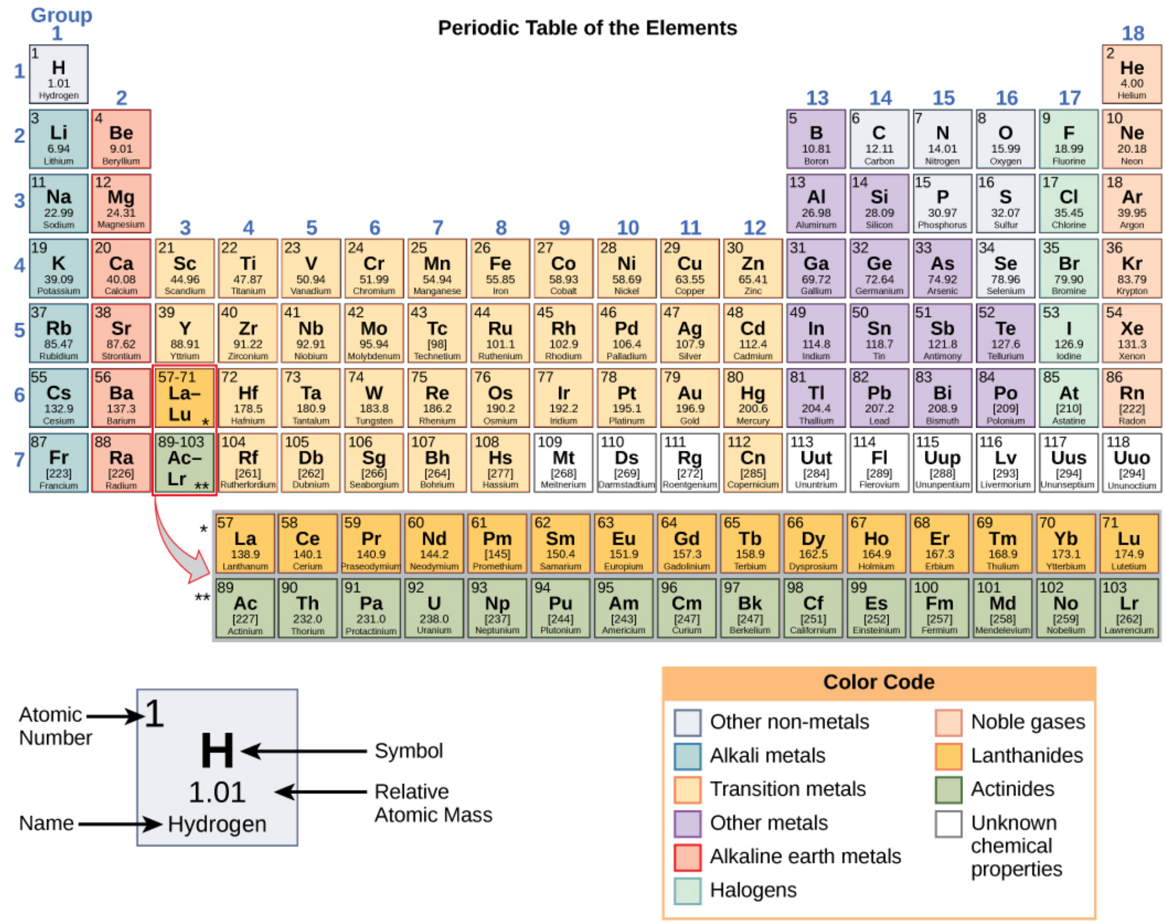
Link to Learning

To learn more about atoms, isotopes, and how to tell one isotope from another, visit this site and run the simulation.

The Periodic Table

The different elements are organized and displayed in the **periodic table**. Devised by Russian chemist Dmitri Mendeleev (1834–1907) in 1869, the table groups elements that, although unique, share certain chemical properties with other elements. The properties of elements are responsible for their physical state at room temperature: they may be gases, solids, or liquids. Elements also have specific **chemical reactivity**, the ability to combine and to chemically bond with each other.

In the periodic table, shown in [link], the elements are organized and displayed according to their atomic number and are arranged in a series of rows and columns based on shared chemical and physical properties. In addition to providing the atomic number for each element, the periodic table also displays the element’s atomic mass. Looking at carbon, for example, its symbol (C) and name appear, as well as its atomic number of six (in the upper left-hand corner) and its atomic mass of 12.11.



Periodic Table of the Elements

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 **H** 1.01 Hydrogen																	2 **He** 4.00 Helium
2	3 **Li** 6.94 Lithium	4 **Be** 9.01 Beryllium											5 **B** 10.81 Boron	6 **C** 12.11 Carbon	7 **N** 14.01 Nitrogen	8 **O** 15.99 Oxygen	9 **F** 18.99 Fluorine	10 **Ne** 20.18 Neon
3	11 **Na** 22.99 Sodium	12 **Mg** 24.31 Magnesium											13 **Al** 26.98 Aluminum	14 **Si** 28.09 Silicon	15 **P** 30.97 Phosphorus	16 **S** 32.07 Sulfur	17 **Cl** 35.45 Chlorine	18 **Ar** 39.95 Argon
4	19 **K** 39.09 Potassium	20 **Ca** 40.08 Calcium	21 **Sc** 44.96 Scandium	22 **Ti** 47.87 Titanium	23 **V** 50.94 Vanadium	24 **Cr** 51.99 Chromium	25 **Mn** 54.94 Manganese	26 **Fe** 55.85 Iron	27 **Co** 58.93 Cobalt	28 **Ni** 58.69 Nickel	29 **Cu** 63.55 Copper	30 **Zn** 65.41 Zinc	31 **Ga** 69.72 Gallium	32 **Ge** 72.64 Germanium	33 **As** 74.92 Arsenic	34 **Se** 78.96 Selenium	35 **Br** 79.90 Bromine	36 **Kr** 83.79 Krypton
5	37 **Rb** 85.47 Rubidium	38 **Sr** 87.62 Strontium	39 **Y** 88.91 Yttrium	40 **Zr** 91.22 Zirconium	41 **Nb** 92.91 Niobium	42 **Mo** 95.94 Molybdenum	43 **Tc** [98] Technetium	44 **Ru** 101.1 Ruthenium	45 **Rh** 102.9 Rhodium	46 **Pd** 106.4 Palladium	47 **Ag** 107.9 Silver	48 **Cd** 112.4 Cadmium	49 **In** 114.8 Indium	50 **Sn** 118.7 Tin	51 **Sb** 121.8 Antimony	52 **Te** 127.6 Tellurium	53 **I** 126.9 Iodine	54 **Xe** 131.3 Xenon
6	55 **Cs** 132.9 Cesium	56 **Ba** 137.3 Barium	57-71 **La–Lu** *	72 **Hf** 178.5 Hafnium	73 **Ta** 180.9 Tantalum	74 **W** 183.8 Tungsten	75 **Re** 186.2 Rhenium	76 **Os** 190.2 Osmium	77 **Ir** 192.2 Iridium	78 **Pt** 195.1 Platinum	79 **Au** 196.9 Gold	80 **Hg** 200.6 Mercury	81 **Tl** 204.4 Thallium	82 **Pb** 207.2 Lead	83 **Bi** 208.9 Bismuth	84 **Po** [209] Polonium	85 **At** [210] Astatine	86 **Rn** [222] Radon
7	87 **Fr** [223] Francium	88 **Ra** [226] Radium	89-103 **Ac–Lr** **	104 **Rf** [261] Rutherfordium	105 **Db** [262] Dubnium	106 **Sg** [266] Seaborgium	107 **Bh** [264] Bohrium	108 **Hs** [277] Hassium	109 **Mt** [268] Meitnerium	110 **Ds** [269] Darmstadtium	111 **Rg** [272] Roentgenium	112 **Cn** [285] Copernicium	113 **Uut** [284] Ununtrium	114 **Fl** [289] Flerovium	115 **Uup** [288] Ununpentium	116 **Lv** [293] Livermorium	117 **Uus** [294] Ununseptium	118 **Uuo** [294] Ununoctium
			*	57 **La** 138.9 Lanthanum	58 **Ce** 140.1 Cerium	59 **Pr** 140.9 Praseodymium	60 **Nd** 144.2 Neodymium	61 **Pm** [145] Promethium	62 **Sm** 150.4 Samarium	63 **Eu** 151.9 Europium	64 **Gd** 157.3 Gadolinium	65 **Tb** 158.9 Terbium	66 **Dy** 162.5 Dysprosium	67 **Ho** 164.9 Holmium	68 **Er** 167.3 Erbium	69 **Tm** 168.9 Thulium	70 **Yb** 173.1 Ytterbium	71 **Lu** 174.9 Lutetium
			**	89 **Ac** [227] Actinium	90 **Th** 232.0 Thorium	91 **Pa** 231.0 Protactinium	92 **U** 238.0 Uranium	93 **Np** [237] Neptunium	94 **Pu** [244] Plutonium	95 **Am** [243] Americium	96 **Cm** [247] Curium	97 **Bk** [247] Berkelium	98 **Cf** [251] Californium	99 **Es** [252] Einsteinium	100 **Fm** [257] Fermium	101 **Md** [258] Mendelevium	102 **No** [259] Nobelium	103 **Lr** [262] Lawrencium

Color Code

☐ Other non-metals	☐ Noble gases
☐ Alkali metals	☐ Lanthanides
☐ Transition metals	☐ Actinides
☐ Other metals	☐ Unknown chemical properties
☐ Alkaline earth metals	
☐ Halogens	

The periodic table shows the atomic mass and atomic number of each element. The atomic number appears above the symbol for the element and the approximate atomic mass appears below it.

The periodic table groups elements according to chemical properties. The differences in chemical reactivity between the elements are based on the number and spatial distribution of an atom’s electrons. Atoms that chemically react and bond to each other form molecules. **Molecules** are simply two or more atoms chemically bonded together. Logically, when two atoms chemically bond to form a molecule, their electrons, which form the outermost region of each atom, come together first as the atoms form a chemical bond.

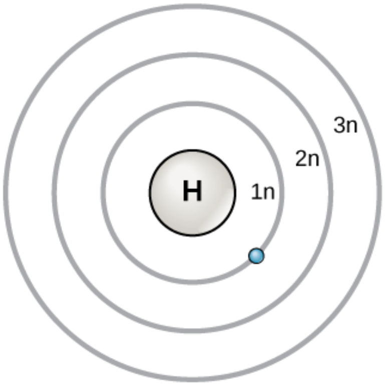
Note:

Visit this website to view the periodic table. In the periodic table of the elements, elements in a single column have the same number of electrons that can participate in a chemical reaction. These electrons are known as “valence electrons.” For example, the elements in the first column all have a single valence electron, an electron that can be “donated” in a chemical reaction with another atom. What is the meaning of a mass number shown in parentheses?

Electron Shells and the Bohr Model

It should be stressed that there is a connection between the number of protons in an element, the atomic number that distinguishes one element from another, and the number of electrons it has. In all electrically neutral atoms, the number of electrons is the same as the number of protons. Thus, each element, at least when electrically neutral, has a characteristic number of electrons equal to its atomic number.

An early model of the atom was developed in 1913 by Danish scientist Niels Bohr (1885–1962). The Bohr model shows the atom as a central nucleus containing protons and neutrons, with the electrons in circular **orbitals** at specific distances from the nucleus, as illustrated in [link]. These orbits form electron shells or energy levels, which are a way of visualizing the number of electrons in the outermost shells. These energy levels are designated by a number and the symbol “n.” For example, 1n represents the first energy level located closest to the nucleus.

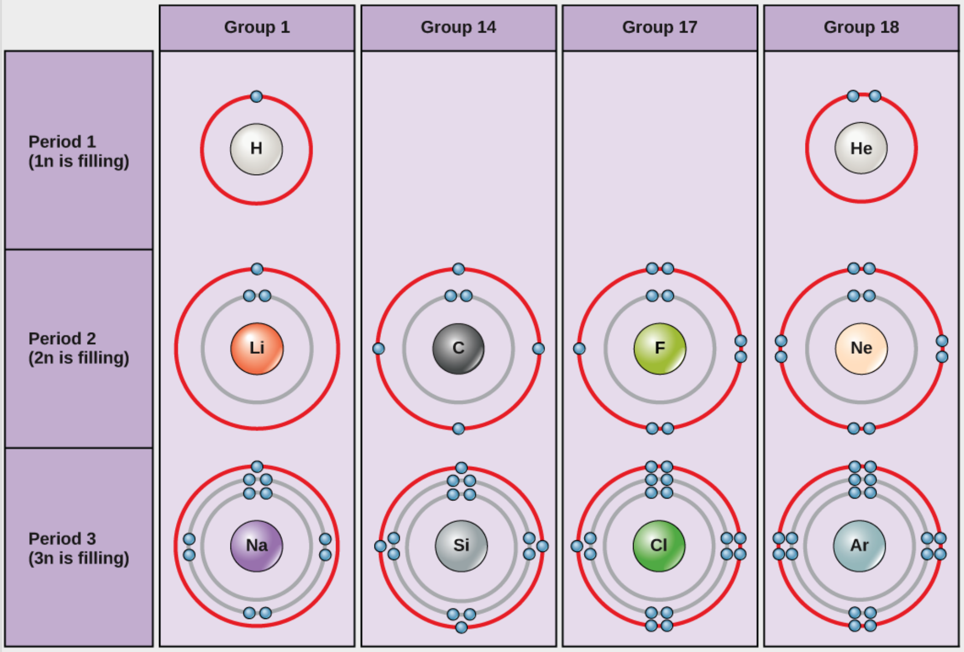


The Bohr model was developed by Niels Bohrs in 1913. In this model, electrons exist within principal shells. An electron normally exists in the lowest energy shell available, which is the one closest to the nucleus. Energy from a photon of light can bump it up to a higher energy shell, but this situation is unstable, and the electron quickly decays back to the ground state. In the process, a photon of light is released.

Electrons fill orbitals in a consistent order: they first fill the orbitals closest to the nucleus, then they continue to fill orbitals of increasing energy further from the nucleus. If there are multiple orbitals of equal energy, they will be filled with one electron in each energy level before a second electron is added. The electrons of the outermost energy level determine the energetic stability of the atom and its tendency to form chemical bonds with other atoms to form molecules.

Under standard conditions, atoms fill the inner shells first, often resulting in a variable number of electrons in the outermost shell. The innermost shell has a maximum of two electrons but the next two electron shells can each have a maximum of eight electrons. This is known as the **octet rule**, which states, with the exception of the innermost shell, that atoms are more stable energetically when they have eight electrons in their **valence shell**, the outermost electron shell. Examples of some neutral atoms and their electron configurations are shown in [link]. Notice that in this [link], helium has a complete outer electron shell, with two electrons filling its first and only shell. Similarly, neon has a complete outer 2n shell containing eight electrons. In contrast, chlorine and sodium have seven and one in their outer shells, respectively, but theoretically they would be more energetically stable if they followed the octet rule and had eight.

Note:
Art Connection



Bohr diagrams indicate how many electrons fill each principal shell. Group 18 elements (helium, neon, and argon are shown) have a full outer, or valence, shell. A full valence shell is the most stable electron configuration. Elements in other groups have partially filled valence shells and gain or lose electrons to achieve a stable electron configuration.

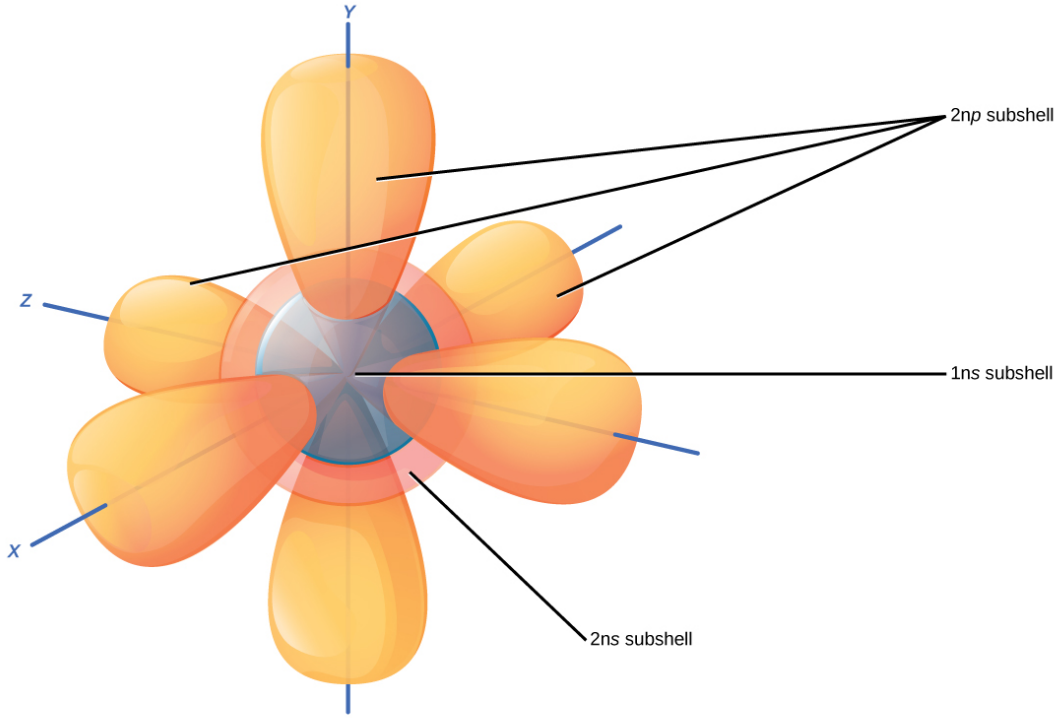
An atom may give, take, or share electrons with another atom to achieve a full valence shell, the most stable electron configuration. Looking at this figure, how many electrons do elements in group 1 need to lose in order to achieve a stable electron configuration? How many electrons do elements in groups 14 and 17 need to gain to achieve a stable configuration?

Understanding that the organization of the periodic table is based on the total number of protons (and electrons) helps us know how electrons are distributed among the outer shell. The periodic table is arranged in columns and rows based on the number of electrons and where these electrons are located. Take a closer look at the some of the elements in the table's far right column in [link]. The group 18 atoms helium (He), neon (Ne), and argon (Ar) all have filled outer electron shells, making it unnecessary for them to share electrons with other atoms to attain stability; they are highly stable as single atoms. Their non-reactivity has resulted in their being named the **inert gases** (or **noble gases**). Compare this to the group 1 elements in the left-hand column. These elements, including hydrogen (H), lithium (Li), and sodium (Na), all have one electron in their outermost shells. That means that they can achieve a stable configuration and a filled outer shell by donating or sharing one electron with another atom or a molecule such as water. Hydrogen will donate or share its electron to achieve this configuration, while lithium and sodium will donate their electron to become stable. As a result of losing a negatively charged electron, they become positively charged **ions**. Group 17 elements, including fluorine and chlorine, have seven electrons in their outmost shells, so they tend to fill this shell with an electron from other atoms or molecules, making them negatively charged ions. Group 14 elements, of which carbon is the most important to living systems, have four electrons in their outer shell allowing them to make several covalent bonds (discussed below) with other atoms. Thus, the columns of the periodic table represent the potential shared state of these elements’ outer electron shells that is responsible for their similar chemical characteristics.

Electron Orbitals

Although useful to explain the reactivity and chemical bonding of certain elements, the Bohr model of the atom does not accurately reflect how electrons are spatially distributed surrounding the nucleus. They do not circle the nucleus like the earth orbits the sun, but are found in **electron orbitals**. These relatively complex shapes result from the fact that electrons behave not just like particles, but also like waves. Mathematical equations from quantum mechanics known as wave functions can predict within a certain level of probability where an electron might be at any given time. The area where an electron is most likely to be found is called its orbital.

Recall that the Bohr model depicts an atom’s electron shell configuration. Within each electron shell are subshells, and each subshell has a specified number of orbitals containing electrons. While it is impossible to calculate exactly where an electron is located, scientists know that it is most probably located within its orbital path. Subshells are designated by the letter s, *p*, *d*, and *f*. The *s* subshell is spherical in shape and has one orbital. Principal shell 1n has only a single *s* orbital, which can hold two electrons. Principal shell 2n has one *s* and one *p* subshell, and can hold a total of eight electrons. The *p* subshell has three dumbbell-shaped orbitals, as illustrated in [link]. Subshells *d* and *f* have more complex shapes and contain five and seven orbitals, respectively. These are not shown in the illustration. Principal shell 3n has *s*, *p*, and *d* subshells and can hold 18 electrons. Principal shell 4n has *s*, *p*, *d* and *f* orbitals and can hold 32 electrons. Moving away from the nucleus, the number of electrons and orbitals found in the energy levels increases. Progressing from one atom to the next in the periodic table, the electron structure can be worked out by fitting an extra electron into the next available orbital.



The *s* subshells are shaped like spheres. Both the 1n and 2n principal shells have an *s* orbital, but the size of the sphere is larger in the 2n orbital. Each sphere is a single orbital. *p* subshells are made up of three dumbbell-shaped orbitals. Principal shell 2n has a *p* subshell, but shell 1 does not.

The closest orbital to the nucleus, called the 1s orbital, can hold up to two electrons. This orbital is equivalent to the innermost electron shell of the Bohr model of the atom. It is called the 1*s* orbital because it is spherical around the nucleus. The 1*s* orbital is the closest orbital to the nucleus, and it is always filled first, before any other orbital can be filled. Hydrogen has one electron; therefore, it has only one spot within the 1*s* orbital occupied. This is designated as $1s^1$, where the superscripted 1 refers to the one electron within the 1*s* orbital. Helium has two electrons; therefore, it can completely fill the 1*s* orbital with its two electrons. This is designated as $1s^2$, referring to the two electrons of helium in the 1*s* orbital. On the periodic table [link], hydrogen and helium are the only two elements in the first row (period); this is because they only have electrons in their first shell, the 1*s* orbital. Hydrogen and helium are the only two elements that have the 1*s* and no other electron orbitals in the electrically neutral state.

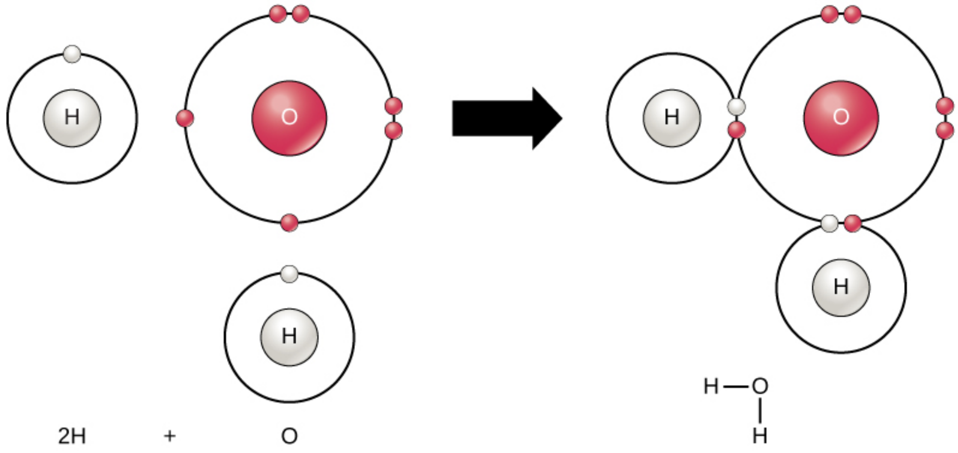
The second electron shell may contain eight electrons. This shell contains another spherical *s* orbital and three “dumbbell” shaped *p* orbitals, each of which can hold two electrons, as shown in [link]. After the 1*s* orbital is filled, the second electron shell is filled, first filling its 2*s* orbital and then its three *p* orbitals. When filling the *p* orbitals, each takes a single electron; once each *p* orbital has an electron, a second may be added. Lithium (Li) contains three electrons that occupy the first and second shells. Two electrons fill the 1*s* orbital, and the third electron then fills the 2*s* orbital. Its **electron configuration** is $1s^{2}2s^{1}$. Neon (Ne), on the other hand, has a total of ten electrons: two are in its innermost 1*s* orbital and eight fill its second shell (two each in the 2*s* and three *p* orbitals); thus, it is an inert gas and energetically stable as a single atom that will rarely form a chemical bond with other atoms. Larger elements have additional orbitals, making up the third electron shell. While the concepts of electron shells and orbitals are closely related, orbitals provide a more accurate depiction of the electron configuration of an atom because the orbital model specifies the different shapes and special orientations of all the places that electrons may occupy.

Note:
Link to Learning

Watch this visual animation to see the spatial arrangement of the *p* and *s* orbitals.

Chemical Reactions and Molecules

All elements are most stable when their outermost shell is filled with electrons according to the octet rule. This is because it is energetically favorable for atoms to be in that configuration and it makes them stable. However, since not all elements have enough electrons to fill their outermost shells, atoms form **chemical bonds** with other atoms thereby obtaining the electrons they need to attain a stable electron configuration. When two or more atoms chemically bond with each other, the resultant chemical structure is a molecule. The familiar water molecule, H_2O, consists of two hydrogen atoms and one oxygen atom; these bond together to form water, as illustrated in [link]. Atoms can form molecules by donating, accepting, or sharing electrons to fill their outer shells.



Two or more atoms may bond with each other to form a molecule. When two hydrogens and an oxygen share electrons via covalent bonds, a water molecule is formed.

Chemical reactions occur when two or more atoms bond together to form molecules or when bonded atoms are broken apart. The substances used in the beginning of a chemical reaction are called the **reactants** (usually found on the left side of a chemical equation), and the substances found at the end of the reaction are known as the **products** (usually found on the right side of a chemical equation). An arrow is typically drawn between the reactants and products to indicate the direction of the chemical reaction; this direction is not always a "one-way street." For the creation of the water molecule shown above, the chemical equation would be:

Equation:

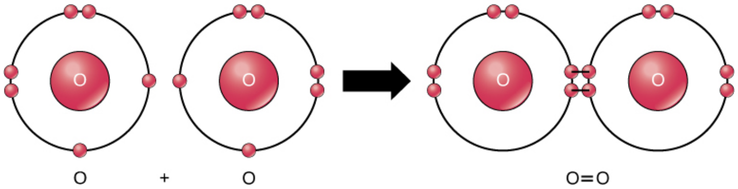
$$2\text{H} + \text{O} \rightarrow \text{H}_2\text{O}$$

An example of a simple chemical reaction is the breaking down of hydrogen peroxide molecules, each of which consists of two hydrogen atoms bonded to two oxygen atoms (H_2O_2). The reactant hydrogen peroxide is broken down into water, containing one oxygen atom bound to two hydrogen atoms (H_2O), and oxygen, which consists of two bonded oxygen atoms (O_2). In the equation below, the reaction includes two hydrogen peroxide molecules and two water molecules. This is an example of a **balanced chemical equation**, wherein the number of atoms of each element is the same on each side of the equation. According to the law of conservation of matter, the number of atoms before and after a chemical reaction should be equal, such that no atoms are, under normal circumstances, created or destroyed.

Equation:

$$2\text{H}_2\text{O}_2\ (\text{hydrogen peroxide}) \rightarrow 2\text{H}_2\text{O}\ (\text{water}) + \text{O}_2\ (\text{oxygen})$$

Even though all of the reactants and products of this reaction are molecules (each atom remains bonded to at least one other atom), in this reaction only hydrogen peroxide and water are representatives of **compounds**: they contain atoms of more than one type of element. Molecular oxygen, on the other hand, as shown in [link],consists of two doubly bonded oxygen atoms and is not classified as a compound but as a mononuclear molecule.



The oxygen atoms in an O_2 molecule are joined by a double bond.

Some chemical reactions, such as the one shown above, can proceed in one direction until the reactants are all used up. The equations that describe these reactions contain a unidirectional arrow and are **irreversible**. **Reversible reactions** are those that can go in either direction. In reversible reactions, reactants are turned into products, but when the concentration of product goes beyond a certain threshold (characteristic of the particular reaction), some of these products will be converted back into reactants; at this point, the designations of products and reactants are reversed. This back and forth continues until a certain relative balance between reactants and products occurs—a state called **equilibrium**. These situations of reversible reactions are often denoted by a chemical equation with a double headed arrow pointing towards both the reactants and products.

For example, in human blood, excess hydrogen ions (H^+) bind to bicarbonate ions (HCO_3^-) forming an equilibrium state with carbonic acid (H_2CO_3). If carbonic acid were added to this system, some of it would be converted to bicarbonate and hydrogen ions.

Equation:

$$HCO_3{}^{-} + H^{+} \leftrightarrow H_2CO_3$$

In biological reactions, however, equilibrium is rarely obtained because the concentrations of the reactants or products or both are constantly changing, often with a product of one reaction being a reactant for another. To return to the example of excess hydrogen ions in the blood, the formation of carbonic acid will be the major direction of the reaction. However, the carbonic acid can also leave the body as carbon dioxide gas (via exhalation) instead of being converted back to bicarbonate ion, thus driving the reaction to the right by the chemical law known as **law of mass action**. These reactions are important for maintaining the homeostasis of our blood.

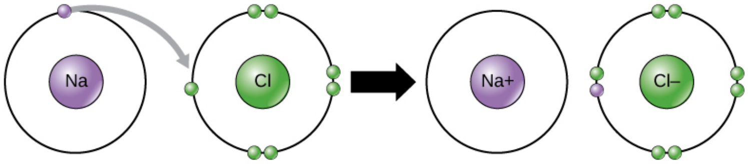
Equation:

$$\mathrm{HCO_3}^- + \mathrm{H}^+ \leftrightarrow \mathrm{H_2CO_3} \leftrightarrow \mathrm{CO_2} + \mathrm{H_2O}$$

Ions and Ionic Bonds

Some atoms are more stable when they gain or lose an electron (or possibly two) and form ions. This fills their outermost electron shell and makes them energetically more stable. Because the number of electrons does not equal the number of protons, each ion has a net charge. **Cations** are positive ions that are formed by losing electrons. Negative ions are formed by gaining electrons and are called anions. **Anions** are designated by their elemental name being altered to end in “-ide”: the anion of chlorine is called chloride, and the anion of sulfur is called sulfide, for example.

This movement of electrons from one element to another is referred to as **electron transfer**. As [link] illustrates, sodium (Na) only has one electron in its outer electron shell. It takes less energy for sodium to donate that one electron than it does to accept seven more electrons to fill the outer shell. If sodium loses an electron, it now has 11 protons, 11 neutrons, and only 10 electrons, leaving it with an overall charge of +1. It is now referred to as a sodium ion. Chlorine (Cl) in its lowest energy state (called the ground state) has seven electrons in its outer shell. Again, it is more energy-efficient for chlorine to gain one electron than to lose seven. Therefore, it tends to gain an electron to create an ion with 17 protons, 17 neutrons, and 18 electrons, giving it a net negative (–1) charge. It is now referred to as a chloride ion. In this example, sodium will donate its one electron to empty its shell, and chlorine will accept that electron to fill its shell. Both ions now satisfy the octet rule and have complete outermost shells. Because the number of electrons is no longer equal to the number of protons, each is now an ion and has a +1 (sodium cation) or –1 (chloride anion) charge. Note that these transactions can normally only take place simultaneously: in order for a sodium atom to lose an electron, it must be in the presence of a suitable recipient like a chlorine atom.



In the formation of an ionic compound, metals lose electrons and nonmetals gain electrons to achieve an octet.

Ionic bonds are formed between ions with opposite charges. For instance, positively charged sodium ions and negatively charged chloride ions bond together to make crystals of sodium chloride, or table salt, creating a crystalline molecule with zero net charge.

Certain salts are referred to in physiology as **electrolytes** (including sodium, potassium, and calcium), ions necessary for nerve impulse conduction, muscle contractions and water balance. Many sports drinks and dietary supplements provide these ions to replace those lost from the body via sweating during exercise.

Covalent Bonds and Other Bonds and Interactions

Another way the octet rule can be satisfied is by the sharing of electrons between atoms to form **covalent bonds**. These bonds are stronger and much more common than ionic bonds in the molecules of living organisms. Covalent bonds are commonly found in carbon-based organic molecules, such as our DNA and proteins. Covalent bonds are also found in inorganic molecules like H_2O, CO_2, and O_2. One, two, or three pairs of electrons may be shared, making single, double, and triple bonds, respectively. The more covalent bonds between two atoms, the stronger their connection. Thus, triple bonds are the strongest.

The strength of different levels of covalent bonding is one of the main reasons living organisms have a difficult time in acquiring nitrogen for use in constructing their molecules, even though molecular nitrogen, N_2, is the most abundant gas in the atmosphere. Molecular nitrogen consists of two nitrogen atoms triple bonded to each other and, as with all molecules, the sharing of these three pairs of electrons between the two nitrogen atoms allows for the filling of their outer electron shells, making the molecule more stable than the individual nitrogen atoms. This strong triple bond makes it difficult for living systems to break apart this nitrogen in order to use it as constituents of proteins and DNA.

The formation of water molecules provides an example of covalent bonding. The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds, as shown in [link]. The electron from the hydrogen splits its time between the incomplete outer shell of the hydrogen atoms and the incomplete outer shell of the oxygen atoms. To completely fill the outer shell of oxygen, which has six electrons in its outer shell but which would be more stable with eight, two electrons (one from each hydrogen atom) are needed: hence the well-known formula H_2O. The electrons are shared between the two elements to fill the outer shell of each, making both elements more stable.

Note:
Link to Learning

View this short video to see an animation of ionic and covalent bonding.

Polar Covalent Bonds

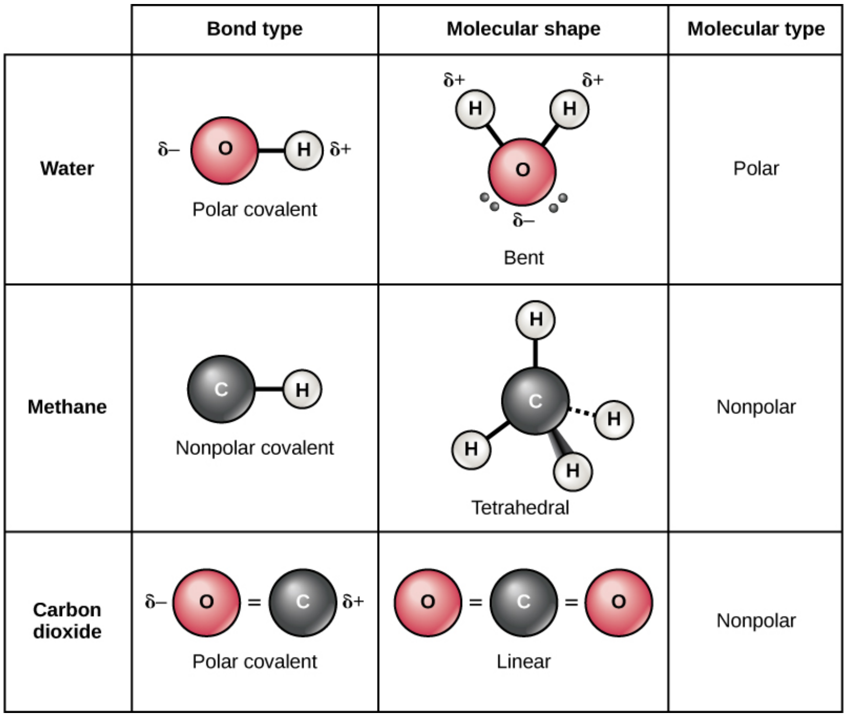
There are two types of covalent bonds: polar and nonpolar. In a **polar covalent bond**, shown in [link], the electrons are unequally shared by the atoms and are attracted more to one nucleus than the other. Because of the unequal distribution of electrons between the atoms of different elements, a slightly positive (δ+) or slightly negative (δ–) charge develops. This partial charge is an important property of water and accounts for many of its characteristics.

Water is a polar molecule, with the hydrogen atoms acquiring a partial positive charge and the oxygen a partial negative charge. This occurs because the nucleus of the oxygen atom is more attractive to the electrons of the hydrogen atoms than the hydrogen nucleus is to the oxygen's electrons. Thus oxygen has a higher **electronegativity** than hydrogen and the shared electrons spend more time in the vicinity of the oxygen nucleus than they do near the nucleus of the hydrogen atoms, giving the atoms of oxygen and hydrogen slightly negative and positive charges, respectively. Another way of stating this is that the probability of finding a shared electron near an oxygen nucleus is more likely than finding it near a hydrogen nucleus. Either way, the atom's relative electronegativity contributes to the development of partial charges whenever one element is significantly more electronegative than the other, and the charges generated by these polar bonds may then be used for the formation of hydrogen bonds based on the attraction of opposite partial charges. (Hydrogen bonds, which are discussed in detail below, are weak bonds between slightly positively charged hydrogen atoms to slightly negatively charged atoms in other molecules.) Since macromolecules often have atoms within them that differ in electronegativity, polar bonds are often present in organic molecules.

Nonpolar Covalent Bonds

Nonpolar covalent bonds form between two atoms of the same element or between different elements that share electrons equally. For example, molecular oxygen (O_2) is nonpolar because the electrons will be equally distributed between the two oxygen atoms.

Another example of a nonpolar covalent bond is methane (CH_4), also shown in [link]. Carbon has four electrons in its outermost shell and needs four more to fill it. It gets these four from four hydrogen atoms, each atom providing one, making a stable outer shell of eight electrons. Carbon and hydrogen do not have the same electronegativity but are similar; thus, nonpolar bonds form. The hydrogen atoms each need one electron for their outermost shell, which is filled when it contains two electrons. These elements share the electrons equally among the carbons and the hydrogen atoms, creating a nonpolar covalent molecule.



	Bond type	Molecular shape	Molecular type
Water	Polar covalent	Bent	Polar
Methane	Nonpolar covalent	Tetrahedral	Nonpolar
Carbon dioxide	Polar covalent	Linear	Nonpolar

Whether a molecule is polar or nonpolar depends both on bond type and molecular shape. Both water and carbon dioxide have polar covalent bonds, but carbon dioxide is linear, so the partial charges on the molecule cancel each other out.

Hydrogen Bonds and Van Der Waals Interactions

Ionic and covalent bonds between elements require energy to break. Ionic bonds are not as strong as covalent, which determines their behavior in biological systems. However, not all bonds are ionic or covalent bonds. Weaker bonds can also form between molecules. Two weak bonds that occur frequently are hydrogen bonds and van der Waals interactions. Without these two types of bonds, life as we know it would not exist. Hydrogen bonds provide many of the critical, life-sustaining properties of water and also stabilize the structures of proteins and DNA, the building block of cells.

When polar covalent bonds containing hydrogen form, the hydrogen in that bond has a slightly positive charge because hydrogen's electron is pulled more strongly toward the other element and away from the hydrogen. Because the hydrogen is slightly positive, it will be attracted to neighboring negative charges. When this happens, a weak interaction occurs between the $\delta^+$of the hydrogen from one molecule and the δ– charge on the more electronegative atoms of another molecule, usually oxygen or nitrogen, or within the same molecule. This interaction is called a **hydrogen bond**. This type of bond is common and occurs regularly between water molecules. Individual hydrogen bonds are weak and easily broken; however, they occur in very large numbers in water and in organic polymers, creating a major force in combination. Hydrogen bonds are also responsible for zipping together the DNA double helix.

Like hydrogen bonds, **van der Waals interactions** are weak attractions or interactions between molecules. Van der Waals attractions can occur between any two or more molecules and are dependent on slight fluctuations of the electron densities, which are not always symmetrical around an atom. For these attractions to happen, the molecules need to be very close to one another. These bonds—along with ionic, covalent, and hydrogen bonds—contribute to the three-dimensional structure of the proteins in our cells that is necessary for their proper function.

Note:

Career Connection

Pharmaceutical Chemist

Pharmaceutical chemists are responsible for the development of new drugs and trying to determine the mode of action of both old and new drugs. They are involved in every step of the drug development process. Drugs can be found in the natural environment or can be synthesized in the laboratory. In many cases, potential drugs found in nature are changed chemically in the laboratory to make them safer and more effective, and sometimes synthetic versions of drugs substitute for the version found in nature.

After the initial discovery or synthesis of a drug, the chemist then develops the drug, perhaps chemically altering it, testing it to see if the drug is toxic, and then designing methods for efficient large-scale production. Then, the process of getting the drug approved for human use begins. In the United States, drug approval is handled by the Food and Drug Administration (FDA) and involves a series of large-scale experiments using human subjects to make sure the drug is not harmful and effectively treats the condition it aims to treat. This process often takes several years and requires the participation of physicians and scientists, in addition to chemists, to complete testing and gain approval.

An example of a drug that was originally discovered in a living organism is Paclitaxel (Taxol), an anti-cancer drug used to treat breast cancer. This drug was discovered in the bark of the pacific yew tree. Another example is aspirin, originally isolated from willow tree bark. Finding drugs often means testing hundreds of samples of plants, fungi, and other forms of life to see if any biologically active compounds are found within them. Sometimes, traditional medicine can give modern medicine clues to where an active compound can be found. For example, the use of willow bark to make medicine has been known for thousands of years, dating back to ancient Egypt. It was not until the late 1800s, however, that the aspirin molecule, known as acetylsalicylic acid, was purified and marketed for human use.

Occasionally, drugs developed for one use are found to have unforeseen effects that allow these drugs to be used in other, unrelated ways. For example, the drug minoxidil (Rogaine) was originally developed to treat high blood pressure. When tested on humans, it was noticed that individuals taking the drug would grow new hair. Eventually the drug was marketed to men and women with baldness to restore lost hair.

The career of the pharmaceutical chemist may involve detective work, experimentation, and drug development, all with the goal of making human beings healthier.

Section Summary

Matter is anything that occupies space and has mass. It is made up of elements. All of the 92 elements that occur naturally have unique qualities that allow them to combine in various ways to create molecules, which in turn combine to form cells, tissues, organ systems, and organisms. Atoms, which consist of protons, neutrons, and electrons, are the smallest units of an element that retain all of the properties of that element. Electrons can be transferred, shared, or cause charge disparities between atoms to create bonds, including ionic, covalent, and hydrogen bonds, as well as van der Waals interactions.

Art Connections

Exercise:

Problem:

[link] How many neutrons do carbon-12 and carbon-13 have, respectively?

Solution:

[link] Carbon-12 has six neutrons. Carbon-13 has seven neutrons.

Exercise:

Problem:

[link] An atom may give, take, or share electrons with another atom to achieve a full valence shell, the most stable electron configuration. Looking at this figure, how many electrons do elements in group 1 need to lose in order to achieve a stable electron configuration? How many electrons do elements in groups 14 and 17 need to gain to achieve a stable configuration?

Solution:

[link] Elements in group 1 need to lose one electron to achieve a stable electron configuration. Elements in groups 14 and 17 need to gain four and one electrons, respectively, to achieve a stable configuration.

Review Questions

Exercise:

Problem:

If xenon has an atomic number of 54 and a mass number of 108, how many neutrons does it have?

a. 54
b. 27
c. 100
d. 108

Solution:

A

Exercise:

Problem:

Atoms that vary in the number of neutrons found in their nuclei are called ________.

a. ions
b. neutrons
c. neutral atoms
d. isotopes

Solution:

D

Exercise:

Problem:

Potassium has an atomic number of 19. What is its electron configuration?

a. shells 1 and 2 are full, and shell 3 has nine electrons
b. shells 1, 2 and 3 are full and shell 4 has three electrons
c. shells 1, 2 and 3 are full and shell 4 has one electron
d. shells 1, 2 and 3 are full and no other electrons are present

---

Solution:

C

Exercise:

Problem: Which type of bond represents a weak chemical bond?

a. hydrogen bond
b. atomic bond
c. covalent bond

d. nonpolar covalent bond

Solution:

A

Free Response

Exercise:

Problem: What makes ionic bonds different from covalent bonds?

Solution:

Ionic bonds are created between ions. The electrons are not shared between the atoms, but rather are associated more with one ion than the other. Ionic bonds are strong bonds, but are weaker than covalent bonds, meaning it takes less energy to break an ionic bond compared with a covalent one.

Exercise:

Problem:

Why are hydrogen bonds and van der Waals interactions necessary for cells?

Solution:

Hydrogen bonds and van der Waals interactions form weak associations between different molecules or within different regions of the same molecule. They provide the structure and shape necessary for proteins and DNA within cells so that they function properly.

Glossary

anion
negative ion that is formed by an atom gaining one or more electrons

atom
the smallest unit of matter that retains all of the chemical properties of an element

atomic mass
calculated mean of the mass number for an element’s isotopes

atomic number
total number of protons in an atom

balanced chemical equation
statement of a chemical reaction with the number of each type of atom equalized for both the products and reactants

cation
positive ion that is formed by an atom losing one or more electrons

chemical bond
interaction between two or more of the same or different atoms that results in the formation of molecules

chemical reaction
process leading to the rearrangement of atoms in molecules

chemical reactivity
the ability to combine and to chemically bond with each other

compound
substance composed of molecules consisting of atoms of at least two different elements

covalent bond
type of strong bond formed between two of the same or different elements; forms when electrons are shared between atoms

electrolyte

ion necessary for nerve impulse conduction, muscle contractions and water balance

electron

negatively charged subatomic particle that resides outside of the nucleus in the electron orbital; lacks functional mass and has a negative charge of –1 unit

electron configuration

arrangement of electrons in an atom’s electron shell (for example, $1s^22s^22p^6$)

electron orbital

how electrons are spatially distributed surrounding the nucleus; the area where an electron is most likely to be found

electron transfer

movement of electrons from one element to another; important in creation of ionic bonds

electronegativity

ability of some elements to attract electrons (often of hydrogen atoms), acquiring partial negative charges in molecules and creating partial positive charges on the hydrogen atoms

element

one of 118 unique substances that cannot be broken down into smaller substances; each element has unique properties and a specified number of protons

equilibrium

steady state of relative reactant and product concentration in reversible chemical reactions in a closed system

hydrogen bond

weak bond between slightly positively charged hydrogen atoms to slightly negatively charged atoms in other molecules

inert gas
(also, noble gas) element with filled outer electron shell that is unreactive with other atoms

ion
atom or chemical group that does not contain equal numbers of protons and electrons

ionic bond
chemical bond that forms between ions with opposite charges (cations and anions)

irreversible chemical reaction
chemical reaction where reactants proceed uni-directionally to form products

isotope
one or more forms of an element that have different numbers of neutrons

law of mass action
chemical law stating that the rate of a reaction is proportional to the concentration of the reacting substances

mass number
total number of protons and neutrons in an atom

matter
anything that has mass and occupies space

molecule
two or more atoms chemically bonded together

neutron
uncharged particle that resides in the nucleus of an atom; has a mass of one amu

noble gas

see inert gas

nonpolar covalent bond
type of covalent bond that forms between atoms when electrons are shared equally between them

nucleus
core of an atom; contains protons and neutrons

octet rule
rule that atoms are most stable when they hold eight electrons in their outermost shells

orbital
region surrounding the nucleus; contains electrons

periodic table
organizational chart of elements indicating the atomic number and atomic mass of each element; provides key information about the properties of the elements

polar covalent bond
type of covalent bond that forms as a result of unequal sharing of electrons, resulting in the creation of slightly positive and slightly negative charged regions of the molecule

product
molecule found on the right side of a chemical equation

proton
positively charged particle that resides in the nucleus of an atom; has a mass of one amu and a charge of +1

radioisotope
isotope that emits radiation composed of subatomic particles to form more stable elements

reactant

molecule found on the left side of a chemical equation

reversible chemical reaction
chemical reaction that functions bi-directionally, where products may turn into reactants if their concentration is great enough

valence shell
outermost shell of an atom

van der Waals interaction
very weak interaction between molecules due to temporary charges attracting atoms that are very close together

Chemical Reactions

By the end of this section, you will be able to:

- Distinguish between kinetic and potential energy, and between exergonic and endergonic chemical reactions
- Identify four forms of energy important in human functioning
- Describe the three basic types of chemical reactions
- Identify several factors influencing the rate of chemical reactions

One characteristic of a living organism is metabolism, which is the sum total of all of the chemical reactions that go on to maintain that organism's health and life. The bonding processes you have learned thus far are anabolic chemical reactions; that is, they form larger molecules from smaller molecules or atoms. But recall that metabolism can proceed in another direction: in catabolic chemical reactions, bonds between components of larger molecules break, releasing smaller molecules or atoms. Both types of reaction involve exchanges not only of matter, but of energy.

The Role of Energy in Chemical Reactions

Chemical reactions require a sufficient amount of energy to cause the matter to collide with enough precision and force that old chemical bonds can be broken and new ones formed. In general, **kinetic energy** is the form of energy powering any type of matter in motion. Imagine you are building a brick wall. The energy it takes to lift and place one brick atop another is kinetic energy—the energy matter possesses because of its motion. Once the wall is in place, it stores potential energy. **Potential energy** is the energy of position, or the energy matter possesses because of the positioning or structure of its components. If the brick wall collapses, the stored potential energy is released as kinetic energy as the bricks fall.

In the human body, potential energy is stored in the bonds between atoms and molecules. **Chemical energy** is the form of potential energy in which energy is stored in chemical bonds. When those bonds are formed, chemical energy is invested, and when they break, chemical energy is released. Notice that chemical energy, like all energy, is neither created nor destroyed; rather, it is converted from one form to another. When you eat an energy bar before heading out the door for a hike, the honey, nuts, and other foods the bar contains are broken down and rearranged by your body into molecules that your muscle cells convert to kinetic energy.

Chemical reactions that release more energy than they absorb are characterized as exergonic. The catabolism of the foods in your energy bar is an example. Some of the chemical energy stored in the bar is absorbed into molecules your body uses for fuel, but some of it is released—for example, as heat. In contrast, chemical reactions that absorb more energy than they release are endergonic. These reactions require energy input, and the resulting molecule stores not only the chemical energy in the original components, but also the energy that fueled the reaction. Because energy is neither created nor destroyed, where does the energy needed for endergonic reactions come from? In many cases, it comes from exergonic reactions.

Forms of Energy Important in Human Functioning

You have already learned that chemical energy is absorbed, stored, and released by chemical bonds. In addition to chemical energy, mechanical, radiant, and electrical energy are important in human functioning.

- Mechanical energy, which is stored in physical systems such as machines, engines, or the human body, directly powers the movement of matter. When you lift a brick into place on a wall, your muscles provide the mechanical energy that moves the brick.
- Radiant energy is energy emitted and transmitted as waves rather than matter. These waves vary in length from long radio waves and microwaves to short gamma waves emitted from decaying atomic nuclei. The full spectrum of radiant energy is referred to as the electromagnetic spectrum. The body uses the ultraviolet energy of sunlight to convert a compound in skin cells to vitamin D, which is essential to human functioning. The human eye evolved to see the wavelengths that comprise the colors of the rainbow, from red to violet, so that range in the spectrum is called “visible light.”
- Electrical energy, supplied by electrolytes in cells and body fluids, contributes to the voltage changes that help transmit impulses in nerve

and muscle cells.

Characteristics of Chemical Reactions

All chemical reactions begin with a **reactant**, the general term for the one or more substances that enter into the reaction. Sodium and chloride ions, for example, are the reactants in the production of table salt. The one or more substances produced by a chemical reaction are called the **product**.

In chemical reactions, the components of the reactants—the elements involved and the number of atoms of each—are all present in the product(s). Similarly, there is nothing present in the products that are not present in the reactants. This is because chemical reactions are governed by the law of conservation of mass, which states that matter cannot be created or destroyed in a chemical reaction.

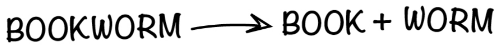
Just as you can express mathematical calculations in equations such as 2 + 7 = 9, you can use chemical equations to show how reactants become products. As in math, chemical equations proceed from left to right, but instead of an equal sign, they employ an arrow or arrows indicating the direction in which the chemical reaction proceeds. For example, the chemical reaction in which one atom of nitrogen and three atoms of hydrogen produce ammonia would be written as $\mathrm{N} + 3\mathrm{H} \rightarrow \mathrm{NH}_3$. Correspondingly, the breakdown of ammonia into its components would be written as $\mathrm{NH}_3 \rightarrow \mathrm{N} + 3\mathrm{H}$.

Notice that, in the first example, a nitrogen (N) atom and three hydrogen (H) atoms bond to form a compound. This anabolic reaction requires energy, which is then stored within the compound's bonds. Such reactions are referred to as synthesis reactions. A **synthesis reaction** is a chemical reaction that results in the synthesis (joining) of components that were formerly separate ([link]**a**). Again, nitrogen and hydrogen are reactants in a synthesis reaction that yields ammonia as the product. The general equation for a synthesis reaction is $\mathrm{A} + \mathrm{B} \rightarrow \mathrm{AB}$.

The Three Fundamental Chemical Reactions

a) In a synthesis reaction, two components bond to make a larger molecule. Energy is required and is stored in the bond:

b) In a decomposition reaction, bonds between components of a larger molecule are broken, resulting in smaller products:



c) In an exchange reaction, bonds are both formed and broken such that the components of the reactants are rearranged:

The atoms and molecules involved in the three fundamental chemical reactions can be imagined as words.

In the second example, ammonia is catabolized into its smaller components, and the potential energy that had been stored in its bonds is released. Such reactions are referred to as decomposition reactions. A **decomposition reaction** is a chemical reaction that breaks down or “de-composes” something larger into its constituent parts (see [link]**b**). The general equation for a decomposition reaction is: $\mathrm{AB} \rightarrow \mathrm{A} + \mathrm{B}$.

An **exchange reaction** is a chemical reaction in which both synthesis and decomposition occur, chemical bonds are both formed and broken, and chemical energy is absorbed, stored, and released (see [link]**c**). The simplest form of an exchange reaction might be: $\mathrm{A} + \mathrm{BC} \rightarrow \mathrm{AB} + \mathrm{C}$. Notice that, to produce these products, B and C had to break apart in a decomposition reaction, whereas A and B had to bond in a synthesis reaction. A more complex exchange reaction might be:$\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{AC} + \mathrm{BD}$. Another example might be: $\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{AD} + \mathrm{BC}$.

In theory, any chemical reaction can proceed in either direction under the right conditions. Reactants may synthesize into a product that is later decomposed. Reversibility is also a quality of exchange reactions. For instance, $\mathrm{A} + \mathrm{BC} \rightarrow \mathrm{AB} + \mathrm{C}$ could then reverse to $\mathrm{AB} + \mathrm{C} \rightarrow \mathrm{A} + \mathrm{BC}$. This reversibility of a chemical reaction is indicated with a double arrow: $\mathrm{A} + \mathrm{BC} \rightleftarrows \mathrm{AB} + \mathrm{C}$. Still, in the human body, many chemical reactions do proceed in a predictable direction, either one way or the other. You can think of this more predictable path as the path of least resistance because, typically, the alternate direction requires more energy.

Oxidation-Reduction Reactions

The chemical reactions underlying metabolism involve the transfer of electrons from one compound to another by processes catalyzed by enzymes. In metabolism the electrons in these reactions commonly, but not always, come from hydrogen atoms, which consist of an electron and a proton. A molecule gives up a hydrogen atom, in the form of a hydrogen ion (H^+) and an electron, breaking the molecule into smaller parts. The loss of an electron, or **oxidation**, releases a small amount of energy; both the electron and the energy are then passed to another molecule in the process of **reduction**, or the gaining of an electron.

The term oxidation is given because oxygen has a high affinity for electrons and is involved in many of these reactions. The result is formation various oxide molecules, for example iron oxide (rust). However, oxygen is only one class of molecule that oxidizes other molecules. Many metabolic functions are driven by oxidizing molecules other than oxygen. Likewise, hydrogen is not the only molecule that is oxidized. For example, redox reactions with iron are important for oxygen transfer in the blood using the iron heme in hemoglobin as a reducing agent.

Like two sides of a coin, these two reactions must always happen as a single **oxidation-reduction reaction** (also called a redox reaction). For one molecule to give up an electron another molecule has to take the electron. The oxidized molecule loses energy in the reaction while the reduced molecule gains energy. Oxidation-reduction reactions often happen in a series, so that a molecule that is reduced is subsequently oxidized, passing on not only the electron it just received but also the energy it received. In this way redox reactions provide the power for cellular activity.

Oxidation-reduction reactions are catalyzed by enzymes that trigger the removal of hydrogen atoms. Coenzymes work with enzymes and accept hydrogen atoms. The two most common coenzymes of oxidation-reduction reactions are **nicotinamide adenine dinucleotide (NAD)** and **flavin**

adenine dinucleotide (FAD). Their respective reduced coenzymes are **NADH** and **$FADH_2$**, which are energy-containing molecules used to transfer energy during the creation of ATP. As the series of redox reactions progress through intermediate molecules, energy accumulates and ultimately is used to combine P_i and ADP to form ATP, the high-energy molecule that the body uses for fuel. The last electron acceptor molecule in this process is oxygen because of its high affinity for electrons.

Factors Influencing the Rate of Chemical Reactions

If you pour vinegar into baking soda, the reaction is instantaneous; the concoction will bubble and fizz. But many chemical reactions take time. A variety of factors influence the rate of chemical reactions. This section, however, will consider only the most important in human functioning.

Properties of the Reactants

If chemical reactions are to occur quickly, the atoms in the reactants have to have easy access to one another. Thus, the greater the surface area of the reactants, the more readily they will interact. When you pop a cube of cheese into your mouth, you chew it before you swallow it. Among other things, chewing increases the surface area of the food so that digestive chemicals can more easily get at it. As a general rule, gases tend to react faster than liquids or solids, again because it takes energy to separate particles of a substance, and gases by definition already have space between their particles. Similarly, the larger the molecule, the greater the number of total bonds, so reactions involving smaller molecules, with fewer total bonds, would be expected to proceed faster.

In addition, recall that some elements are more reactive than others. Reactions that involve highly reactive elements like hydrogen proceed more quickly than reactions that involve less reactive elements. Reactions involving stable elements like helium are not likely to happen at all.

Temperature

Nearly all chemical reactions occur at a faster rate at higher temperatures. Recall that kinetic energy is the energy of matter in motion. The kinetic energy of subatomic particles increases in response to increases in thermal energy. The higher the temperature, the faster the particles move, and the more likely they are to come in contact and react.

Concentration and Pressure

If just a few people are dancing at a club, they are unlikely to step on each other's toes. But as more and more people get up to dance—especially if the music is fast—collisions are likely to occur. It is the same with chemical reactions: the more particles present within a given space, the more likely those particles are to bump into one another. This means that chemists can speed up chemical reactions not only by increasing the **concentration** of particles—the number of particles in the space—but also by decreasing the volume of the space, which would correspondingly increase the pressure. If there were 100 dancers in that club, and the manager abruptly moved the party to a room half the size, the concentration of the dancers would double in the new space, and the likelihood of collisions would increase accordingly.

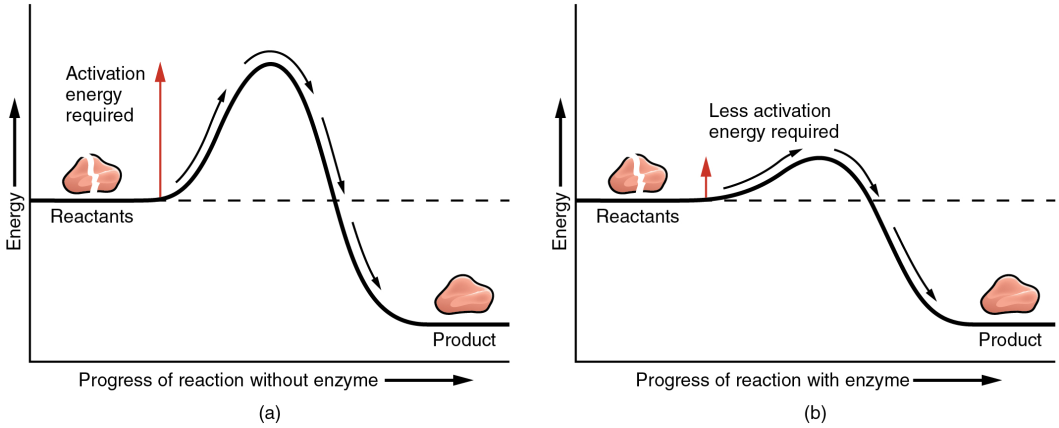
Enzymes and Other Catalysts

For two chemicals in nature to react with each other they first have to come into contact, and this occurs through random collisions. Because heat helps increase the kinetic energy of atoms, ions, and molecules, it promotes their collision. But in the body, extremely high heat—such as a very high fever—can damage body cells and be life-threatening. On the other hand, normal body temperature is not high enough to promote the chemical reactions that sustain life. That is where catalysts come in.

In chemistry, a **catalyst** is a substance that increases the rate of a chemical reaction without itself undergoing any change. You can think of a catalyst as a chemical change agent. They help increase the rate and force at which atoms, ions, and molecules collide, thereby increasing the probability that their valence shell electrons will interact.

The most important catalysts in the human body are enzymes. An **enzyme** is a catalyst composed of protein or ribonucleic acid (RNA), both of which will be discussed later in this chapter. Like all catalysts, enzymes work by lowering the level of energy that needs to be invested in a chemical reaction. A chemical reaction's **activation energy** is the "threshold" level of energy needed to break the bonds in the reactants. Once those bonds are broken, new arrangements can form. Without an enzyme to act as a catalyst, a much larger investment of energy is needed to ignite a chemical reaction ([link]).

Enzymes



Activation energy required
Energy
Reactants
Product
Progress of reaction without enzyme
(a)
Less activation energy required
Energy
Reactants
Product
Progress of reaction with enzyme
(b)

Enzymes decrease the activation energy required for a given chemical reaction to occur. (a) Without an enzyme, the energy input needed for a reaction to begin is high. (b) With the help of an enzyme, less energy is needed for a reaction to begin.

Enzymes are critical to the body's healthy functioning. They assist, for example, with the breakdown of food and its conversion to energy. In fact, most of the chemical reactions in the body are facilitated by enzymes.

Chapter Review

Chemical reactions, in which chemical bonds are broken and formed, require an initial investment of energy. Kinetic energy, the energy of matter in motion, fuels the collisions of atoms, ions, and molecules that are necessary if their old bonds are to break and new ones to form. All molecules store potential energy, which is released when their bonds are broken.

Four forms of energy essential to human functioning are: chemical energy, which is stored and released as chemical bonds are formed and broken; mechanical energy, which directly powers physical activity; radiant energy, emitted as waves such as in sunlight; and electrical energy, the power of moving electrons.

Chemical reactions begin with reactants and end with products. Synthesis reactions bond reactants together, a process that requires energy, whereas decomposition reactions break the bonds within a reactant and thereby release energy. In exchange reactions, bonds are both broken and formed, and energy is exchanged.

The rate at which chemical reactions occur is influenced by several properties of the reactants: temperature, concentration and pressure, and the presence or absence of a catalyst. An enzyme is a catalytic protein that speeds up chemical reactions in the human body.

Review Questions

Exercise:

Problem:

The energy stored in a foot of snow on a steep roof is ________.

a. potential energy
b. kinetic energy
c. radiant energy
d. activation energy

Solution:

A

Exercise:

Problem:

The bonding of calcium, phosphorus, and other elements produces mineral crystals that are found in bone. This is an example of a(n) ________ reaction.

a. catabolic
b. synthesis
c. decomposition
d. exchange

Solution:

B

Exercise:

Problem:

$\mathrm{AB} \rightarrow \mathrm{A} + \mathrm{B}$ is a general notation for a(n) ________ reaction.

a. anabolic
b. endergonic
c. decomposition
d. exchange

Solution:

C

Exercise:

Problem:________ reactions release energy.

a. Catabolic
b. Exergonic
c. Decomposition
d. Catabolic, exergonic, and decomposition

Solution:

D

Exercise:

Problem:

Which of the following combinations of atoms is *most likely* to result in a chemical reaction?

a. hydrogen and hydrogen
b. hydrogen and helium
c. helium and helium
d. neon and helium

Solution:

A

Exercise:

Problem:

Chewing a bite of bread mixes it with saliva and facilitates its chemical breakdown. This is *most likely* due to the fact that ________.

a. the inside of the mouth maintains a very high temperature
b. chewing stores potential energy
c. chewing facilitates synthesis reactions

d. saliva contains enzymes

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

$\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{AD} + \mathrm{BE}$ Is this a legitimate example of an exchange reaction? Why or why not?

Solution:

It is not. An exchange reaction might be $\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{AC} + \mathrm{BD}$ or $\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{AD} + \mathrm{BC}$. In all chemical reactions, including exchange reactions, the components of the reactants are identical to the components of the products. A component present among the reactants cannot disappear, nor can a component not present in the reactants suddenly appear in the products.

Exercise:

Problem:

When you do a load of laundry, why do you not just drop a bar of soap into the washing machine? In other words, why is laundry detergent sold as a liquid or powder?

Solution:

Recall that the greater the surface area of the reactants, the more quickly and easily they will interact. It takes energy to separate particles of a substance. Powder and liquid laundry detergents, with relatively more surface area per unit, can quickly dissolve into their reactive components when added to the water.

Glossary

activation energy
amount of energy greater than the energy contained in the reactants, which must be overcome for a reaction to proceed

catalyst
substance that increases the rate of a chemical reaction without itself being changed in the process

chemical energy
form of energy that is absorbed as chemical bonds form, stored as they are maintained, and released as they are broken

concentration
number of particles within a given space

decomposition reaction
type of catabolic reaction in which one or more bonds within a larger molecule are broken, resulting in the release of smaller molecules or atoms

enzyme
protein or RNA that catalyzes chemical reactions

exchange reaction
type of chemical reaction in which bonds are both formed and broken, resulting in the transfer of components

kinetic energy
energy that matter possesses because of its motion

potential energy

stored energy matter possesses because of the positioning or structure of its components

product
one or more substances produced by a chemical reaction

reactant
one or more substances that enter into the reaction

synthesis reaction
type of anabolic reaction in which two or more atoms or molecules bond, resulting in the formation of a larger molecule

Inorganic Compounds Essential to Human Functioning

By the end of this section, you will be able to:

- Compare and contrast inorganic and organic compounds
- Identify the properties of water that make it essential to life
- Explain the role of salts in body functioning
- Distinguish between acids and bases, and explain their role in pH
- Discuss the role of buffers in helping the body maintain pH homeostasis

The concepts you have learned so far in this chapter govern all forms of matter, and would work as a foundation for geology as well as biology. This section of the chapter narrows the focus to the chemistry of human life; that is, the compounds important for the body’s structure and function. In general, these compounds are either inorganic or organic.

- An **inorganic compound** is a substance that does not contain both carbon and hydrogen. A great many inorganic compounds do contain hydrogen atoms, such as water (H_2O) and the hydrochloric acid (HCl) produced by your stomach. In contrast, only a handful of inorganic compounds contain carbon atoms. Carbon dioxide (CO_2) is one of the few examples.
- An **organic compound**, then, is a substance that contains both carbon and hydrogen. Organic compounds are synthesized via covalent bonds within living organisms, including the human body. Recall that carbon and hydrogen are the second and third most abundant elements in your body. You will soon discover how these two elements combine in the foods you eat, in the compounds that make up your body structure, and in the chemicals that fuel your functioning.

The following section examines the three groups of inorganic compounds essential to life: water, salts, acids, and bases. Organic compounds are covered later in the chapter.

Water

As much as 70 percent of an adult’s body weight is water. This water is contained both within the cells and between the cells that make up tissues and organs. Its several roles make water indispensable to human functioning.

Water as a Lubricant and Cushion

Water is a major component of many of the body’s lubricating fluids. Just as oil lubricates the hinge on a door, water in synovial fluid lubricates the actions of body joints, and water in pleural fluid helps the lungs expand and recoil with breathing. Watery fluids help keep food flowing through the digestive tract, and ensure that the movement of adjacent abdominal organs is friction free.

Water also protects cells and organs from physical trauma, cushioning the brain within the skull, for example, and protecting the delicate nerve tissue of the eyes. Water cushions a developing fetus in the mother’s womb as well.

Water as a Heat Sink

A heat sink is a substance or object that absorbs and dissipates heat but does not experience a corresponding increase in temperature. In the body, water absorbs the heat generated by chemical reactions without greatly increasing in temperature. Moreover, when the environmental temperature soars, the water stored in the body helps keep the body cool. This cooling effect happens as warm blood from the body’s core flows to the blood vessels just under the skin and is transferred to the environment. At the same time, sweat glands release warm water in sweat. As the water evaporates into the air, it carries away heat, and then the cooler blood from the periphery circulates back to the body core.

Water as a Component of Liquid Mixtures

A mixture is a combination of two or more substances, each of which maintains its own chemical identity. In other words, the constituent substances are not chemically bonded into a new, larger chemical compound. The concept is easy to imagine if you think of powdery substances such as flour and sugar; when you stir them together in a bowl, they obviously do not bond to form a new compound. The room air you breathe is a gaseous mixture, containing three discrete elements—nitrogen, oxygen, and argon—and one compound, carbon dioxide. There are three types of liquid mixtures, all of which contain water as a key component. These are solutions, colloids, and suspensions.

For cells in the body to survive, they must be kept moist in a water-based liquid called a solution. In chemistry, a liquid **solution** consists of a solvent that dissolves a substance called a solute. An important characteristic of solutions is that they are homogeneous; that is, the solute molecules are distributed evenly throughout the solution. If you were to stir a teaspoon of sugar into a glass of water, the sugar would dissolve into sugar molecules separated by water molecules. The ratio of sugar to water in the left side of the glass would be the same as the ratio of sugar to water in the right side of the glass. If you were to add more sugar, the ratio of sugar to water would change, but the distribution—provided you had stirred well—would still be even.

Water is considered the “universal solvent” and it is believed that life cannot exist without water because of this. Water is certainly the most abundant solvent in the body; essentially all of the body’s chemical reactions occur among compounds dissolved in water. Because water molecules are polar, with regions of positive and negative electrical charge, water readily dissolves ionic compounds and polar covalent compounds. Such compounds are referred to as hydrophilic, or “water-loving.” As mentioned above, sugar dissolves well in water. This is because sugar molecules contain regions of hydrogen-oxygen polar bonds, making it hydrophilic. Nonpolar molecules, which do not readily dissolve in water, are called hydrophobic, or “water-fearing.”

Concentrations of Solutes

Various mixtures of solutes and water are described in chemistry. The concentration of a given solute is the number of particles of that solute in a given space (oxygen makes up about 21 percent of atmospheric air). In the bloodstream of humans, glucose concentration is usually measured in milligram (mg) per deciliter (dL), and in a healthy adult averages about 100 mg/dL. Another method of measuring the concentration of a solute is by its molarilty—which is moles (M) of the molecules per liter (L). The mole of an element is its atomic weight, while a mole of a compound is the sum of the atomic weights of its components, called the molecular weight. An often-used example is calculating a mole of glucose, with the chemical formula $C_6H_{12}O_6$. Using the periodic table, the atomic weight of carbon (C) is 12.011 grams (g), and there are six carbons in glucose, for a total atomic weight of 72.066 g. Doing the same calculations for hydrogen (H) and oxygen (O), the molecular weight equals 180.156g (the "gram molecular weight" of glucose). When water is added to make one liter of solution, you have one mole (1M) of glucose. This is particularly useful in chemistry because of the relationship of moles to "Avogadro's number." A mole of any solution has the same number of particles in it: 6.02×10^{23}. Many substances in the bloodstream and other tissue of the body are measured in thousandths of a mole, or millimoles (mM).

A **colloid** is a mixture that is somewhat like a heavy solution. The solute particles consist of tiny clumps of molecules large enough to make the liquid mixture opaque (because the particles are large enough to scatter light). Familiar examples of colloids are milk and cream. In the thyroid glands, the thyroid hormone is stored as a thick protein mixture also called a colloid.

A **suspension** is a liquid mixture in which a heavier substance is suspended temporarily in a liquid, but over time, settles out. This separation of particles from a suspension is called sedimentation. An example of sedimentation occurs in the blood test that establishes sedimentation rate, or sed rate. The test measures how quickly red blood cells in a test tube settle out of the watery portion of blood (known as plasma) over a set period of time. Rapid sedimentation of blood cells does not normally happen in the healthy body, but aspects of certain diseases can cause blood cells to clump together, and these heavy clumps of blood cells settle to the bottom of the test tube more quickly than do normal blood cells.

The Role of Water in Chemical Reactions

Two types of chemical reactions involve the creation or the consumption of water: dehydration synthesis and hydrolysis.

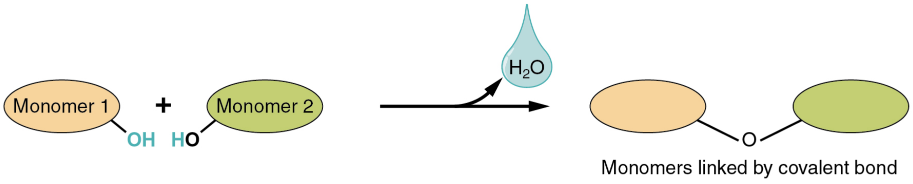
- In dehydration synthesis, one reactant gives up an atom of hydrogen and another reactant gives up a hydroxyl group (OH) in the synthesis of a new product. In the formation of their covalent bond, a molecule of water is released as a byproduct ([link]). This is also sometimes referred to as a condensation reaction.
- In hydrolysis, a molecule of water disrupts a compound, breaking its bonds. The water is itself split into H and OH. One portion of the severed compound then bonds with the hydrogen atom, and the other portion bonds with the hydroxyl group.

These reactions are reversible, and play an important role in the chemistry of organic compounds (which will be discussed shortly).

Dehydration Synthesis and Hydrolysis

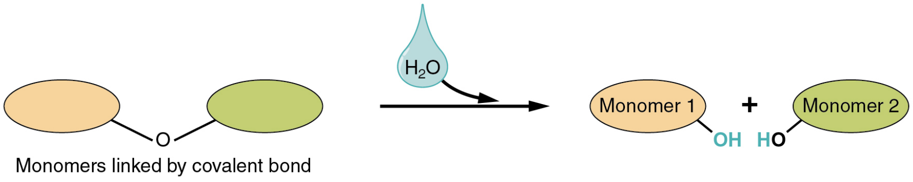
(a) Dehydration synthesis

Monomers are joined by removal of OH from one monomer and removal of H from the other at the site of bond formation.



(b) Hydrolysis

Monomers are released by the addition of a water molecule, adding OH to one monomer and H to the other.



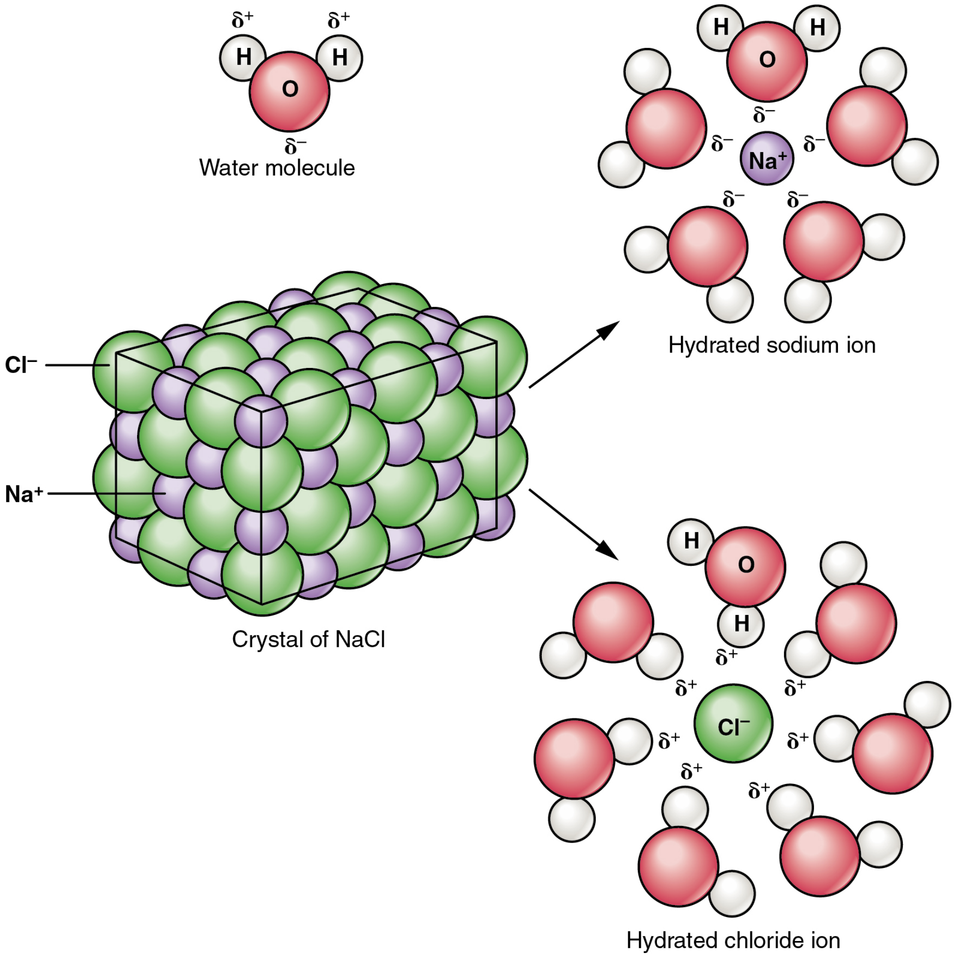
Monomers, the basic units for building larger molecules, form polymers (two or more chemically-bonded monomers). (a) In dehydration synthesis, two monomers are covalently bonded in a reaction in which one gives up a hydroxyl group and the other a hydrogen atom. A molecule of water is released as a byproduct during dehydration reactions. (b) In hydrolysis, the covalent bond between two monomers is split by the addition of a hydrogen atom to one and a hydroxyl group to the other, which requires the contribution of one molecule of water.

Salts

Recall that salts are formed when ions form ionic bonds. In these reactions, one atom gives up one or more electrons, and thus becomes positively charged, whereas the other accepts one or more electrons and becomes negatively charged. You can now define a salt as a substance that, when dissolved in water, dissociates into ions other than H^+ or OH^-. This fact is important in distinguishing salts from acids and bases, discussed next.

A typical salt, NaCl, dissociates completely in water ([link]). The positive and negative regions on the water molecule (the hydrogen and oxygen ends respectively) attract the negative chloride and positive sodium ions, pulling them away from each other. Again, whereas nonpolar and polar covalently bonded compounds break apart into molecules in solution, salts dissociate into ions. These ions are electrolytes; they are capable of conducting an electrical current in solution. This property is critical to the function of ions in transmitting nerve impulses and prompting muscle contraction.

Dissociation of Sodium Chloride in Water



Notice that the crystals of sodium chloride dissociate not into molecules of NaCl, but into Na$^+$ cations and Cl$^-$ anions, each completely surrounded by water molecules.

Many other salts are important in the body. For example, bile salts produced by the liver help break apart dietary fats, and calcium phosphate salts form the mineral portion of teeth and bones.

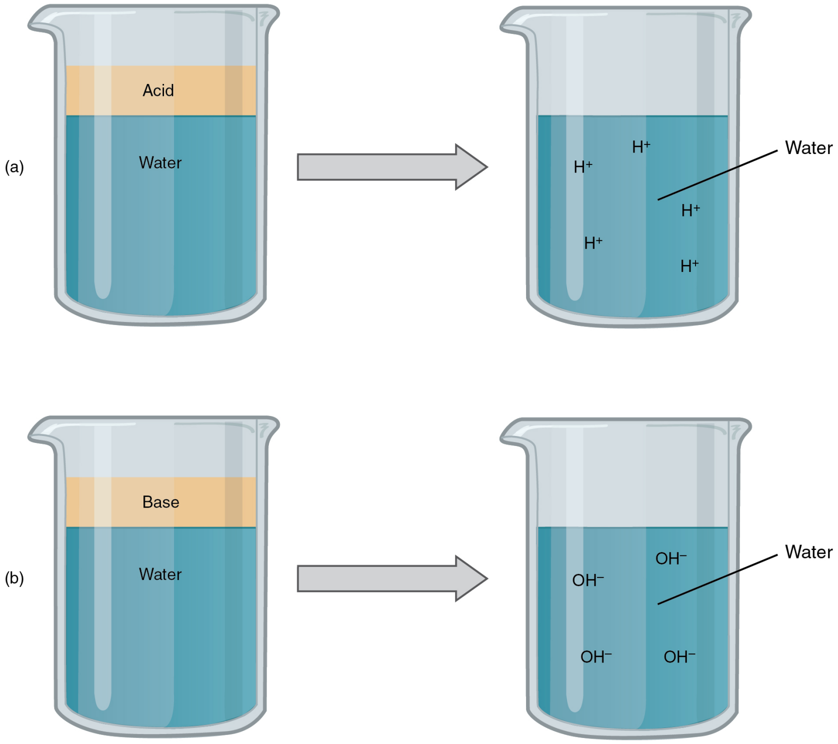
Acids and Bases

Acids and bases, like salts, dissociate in water into electrolytes. Acids and bases can very much change the properties of the solutions in which they are dissolved.

Acids

An **acid** is a substance that releases hydrogen ions (H^+) in solution ([link]**a**). Because an atom of hydrogen has just one proton and one electron, a positively charged hydrogen ion is simply a proton. This solitary proton is highly likely to participate in chemical reactions. Strong acids are compounds that release all of their H^+ in solution; that is, they ionize completely. Hydrochloric acid (HCl), which is released from cells in the lining of the stomach, is a strong acid because it releases all of its H^+ in the stomach's watery environment. This strong acid aids in digestion and kills ingested microbes. Weak acids do not ionize completely; that is, some of their hydrogen ions remain bonded within a compound in solution. An example of a weak acid is vinegar, or acetic acid; it is called acetate after it gives up a proton.

Acids and Bases



(a) In aqueous solution, an acid dissociates into hydrogen ions (H^+) and anions. Nearly every molecule of a strong acid dissociates, producing a high concentration of H^+. (b) In aqueous solution, a base dissociates into hydroxyl ions ($OH^–$) and cations. Nearly every molecule of a strong base dissociates, producing a high concentration of $OH^–$.

Bases

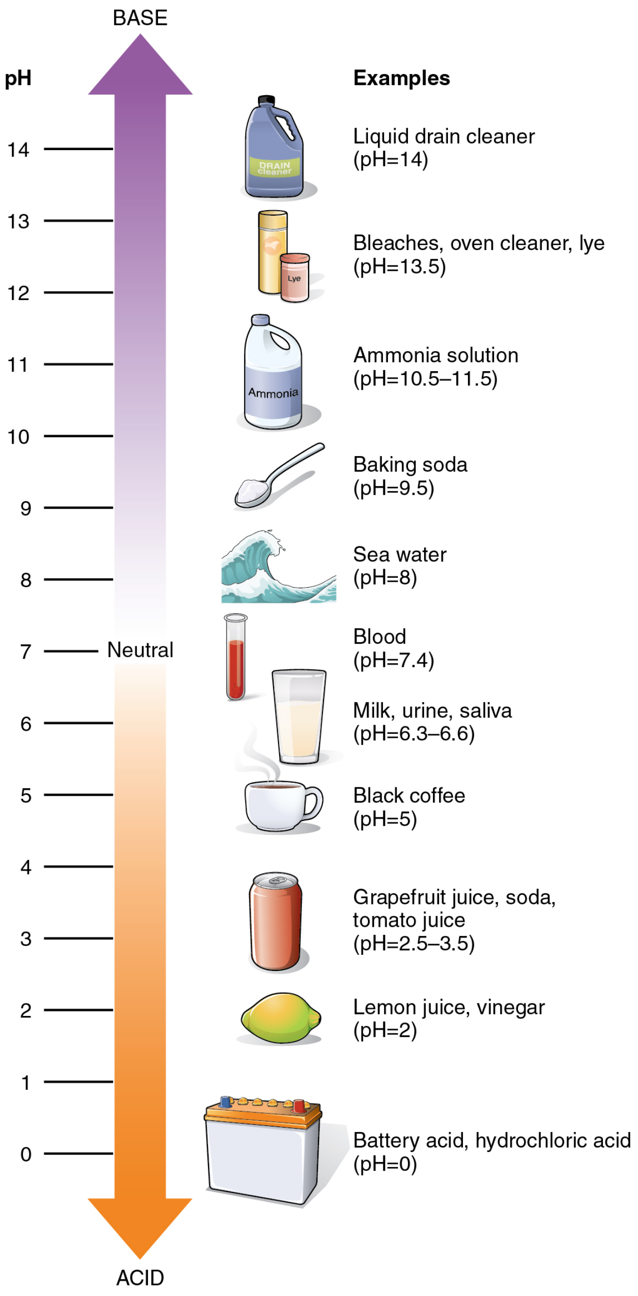
A **base** is a substance that releases hydroxyl ions ($OH^–$) in solution, or one that accepts H^+ already present in solution (see [link]**b**). The hydroxyl ions (also known as hydroxide ions) or other basic substances combine with H^+ present to form a water molecule, thereby removing H^+ and reducing the solution's acidity. Strong bases release most or all of their hydroxyl ions; weak bases release only some hydroxyl ions or absorb only a few H^+. Food mixed with hydrochloric acid from the stomach would burn the small intestine, the next portion of the digestive tract after the stomach, if it were not for the release of bicarbonate ($HCO_3^–$), a weak base that attracts H^+. Bicarbonate accepts some of the H^+ protons, thereby reducing the acidity of the solution.

The Concept of pH

The relative acidity or alkalinity of a solution can be indicated by its pH. A solution's **pH** is the negative, base-10 logarithm of the hydrogen ion (H^+) concentration of the solution. As an example, a pH 4 solution has an H^+ concentration that is ten times greater than that of a pH 5 solution. That is, a solution with a pH of 4 is ten times more acidic than a solution with a pH of 5. The concept of pH will begin to make more sense when you study the pH scale, like that shown in [link]. The scale consists of a series of increments ranging from 0 to 14. A solution with a pH of 7 is considered neutral—

neither acidic nor basic. Pure water has a pH of 7. The lower the number below 7, the more acidic the solution, or the greater the concentration of H^+. The concentration of hydrogen ions at each pH value is 10 times different than the next pH. For instance, a pH value of 4 corresponds to a proton concentration of 10^{-4} M, or 0.0001M, while a pH value of 5 corresponds to a proton concentration of 10^{-5} M, or 0.00001M. The higher the number above 7, the more basic (alkaline) the solution, or the lower the concentration of H^+. Human urine, for example, is ten times more acidic than pure water, and HCl is 10,000,000 times more acidic than water.

The pH Scale



Buffers

The pH of human blood normally ranges from 7.35 to 7.45, although it is typically identified as pH 7.4. At this slightly basic pH, blood can reduce the acidity resulting from the carbon dioxide (CO_2) constantly being released into the bloodstream by the trillions of cells in the body.

Homeostatic mechanisms (along with exhaling CO_2 while breathing) normally keep the pH of blood within this narrow range. This is critical, because fluctuations—either too acidic or too alkaline—can lead to life-threatening disorders.

All cells of the body depend on homeostatic regulation of acid–base balance at a pH of approximately 7.4. The body therefore has several mechanisms for this regulation, involving breathing, the excretion of chemicals in urine, and the internal release of chemicals collectively called buffers into body fluids. A **buffer** is a solution of a weak acid and its conjugate base. A buffer can neutralize small amounts of acids or bases in body fluids. For example, if there is even a slight decrease below 7.35 in the pH of a bodily fluid, the buffer in the fluid—in this case, acting as a weak base—will bind the excess hydrogen ions. In contrast, if pH rises above 7.45, the buffer will act as a weak acid and contribute hydrogen ions.

Note:

Homeostatic Imbalances

Acids and Bases

Excessive acidity of the blood and other body fluids is known as acidosis. Common causes of acidosis are situations and disorders that reduce the effectiveness of breathing, especially the person’s ability to exhale fully, which causes a buildup of CO_2 (and H^+) in the bloodstream. Acidosis can also be caused by metabolic problems that reduce the level or function of buffers that act as bases, or that promote the production of acids. For instance, with severe diarrhea, too much bicarbonate can be lost from the body, allowing acids to build up in body fluids. In people with poorly managed diabetes (ineffective regulation of blood sugar), acids called ketones are produced as a form of body fuel. These can build up in the blood, causing a serious condition called diabetic ketoacidosis. Kidney failure, liver failure, heart failure, cancer, and other disorders also can prompt metabolic acidosis.

In contrast, alkalosis is a condition in which the blood and other body fluids are too alkaline (basic). As with acidosis, respiratory disorders are a major cause; however, in respiratory alkalosis, carbon dioxide levels fall too low. Lung disease, aspirin overdose, shock, and ordinary anxiety can cause respiratory alkalosis, which reduces the normal concentration of H^+. Metabolic alkalosis often results from prolonged, severe vomiting, which causes a loss of hydrogen and chloride ions (as components of HCl). Medications also can prompt alkalosis. These include diuretics that cause the body to lose potassium ions, as well as antacids when taken in excessive amounts, for instance by someone with persistent heartburn or an ulcer.

Chapter Review

Inorganic compounds essential to human functioning include water, salts, acids, and bases. These compounds are inorganic; that is, they do not contain both hydrogen and carbon. Water is a lubricant and cushion, a heat sink, a component of liquid mixtures, a byproduct of dehydration synthesis reactions, and a reactant in hydrolysis reactions. Salts are compounds that, when dissolved in water, dissociate into ions other than H^+ or $OH^–$. In contrast, acids release H^+ in solution, making it more acidic. Bases accept H^+, thereby making the solution more alkaline (caustic).

The pH of any solution is its relative concentration of H^+. A solution with pH 7 is neutral. Solutions with pH below 7 are acids, and solutions with pH above 7 are bases. A change in a single digit on the pH scale (e.g., from 7 to 8) represents a ten-fold increase or decrease in the concentration of H^+. In a healthy adult, the pH of blood ranges from 7.35 to 7.45. Homeostatic control mechanisms important for keeping blood in a healthy pH range include chemicals called buffers, weak acids and weak bases released when the pH of blood or other body fluids fluctuates in either direction outside of this normal range.

Review Questions

Exercise:

Problem:CH_4 is methane. This compound is ________.

a. inorganic
b. organic
c. reactive
d. a crystal

Solution:

B

Exercise:

Problem:

Which of the following is most likely to be found evenly distributed in water in a homogeneous solution?

a. sodium ions and chloride ions
b. NaCl molecules
c. salt crystals
d. red blood cells

Solution:

A

Exercise:

Problem:

Jenny mixes up a batch of pancake batter, then stirs in some chocolate chips. As she is waiting for the first few pancakes to cook, she notices the chocolate chips sinking to the bottom of the clear glass mixing bowl. The chocolate-chip batter is an example of a ________.

a. solvent
b. solute
c. solution
d. suspension

Solution:

D

Exercise:

Problem:

A substance dissociates into K^+ and Cl^- in solution. The substance is a(n) ________.

a. acid
b. base
c. salt
d. buffer

Solution:

C

Exercise:

Problem:

Ty is three years old and as a result of a “stomach bug” has been vomiting for about 24 hours. His blood pH is 7.48. What does this mean?

a. Ty’s blood is slightly acidic.
b. Ty’s blood is slightly alkaline.
c. Ty’s blood is highly acidic.
d. Ty’s blood is within the normal range

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

The pH of lemon juice is 2, and the pH of orange juice is 4. Which of these is more acidic, and by how much? What does this mean?

Solution:

Lemon juice is one hundred times more acidic than orange juice. This means that lemon juice has a one hundred-fold greater concentration of hydrogen ions.

Exercise:

Problem:

During a party, Eli loses a bet and is forced to drink a bottle of lemon juice. Not long thereafter, he begins complaining of having difficulty breathing, and his friends take him to the local emergency room. There, he is given an intravenous solution of bicarbonate. Why?

Solution:

Lemon juice, like any acid, releases hydrogen ions in solution. As excessive H^+ enters the digestive tract and is absorbed into blood, Eli's blood pH falls below 7.35. Recall that bicarbonate is a buffer, a weak base that accepts hydrogen ions. By administering bicarbonate intravenously, the emergency department physician helps raise Eli's blood pH back toward neutral.

Glossary

acid
compound that releases hydrogen ions (H^+) in solution

base

compound that accepts hydrogen ions (H^+) in solution

buffer
solution containing a weak acid or a weak base that opposes wide fluctuations in the pH of body fluids

colloid
liquid mixture in which the solute particles consist of clumps of molecules large enough to scatter light

inorganic compound
substance that does not contain both carbon and hydrogen

organic compound
substance that contains both carbon and hydrogen

pH
negative logarithm of the hydrogen ion (H^+) concentration of a solution

solution
homogeneous liquid mixture in which a solute is dissolved into molecules within a solvent

suspension
liquid mixture in which particles distributed in the liquid settle out over time

Organic Compounds Essential to Human Functioning
By the end of this section, you will be able to:

- Identify four types of organic molecules essential to human functioning
- Explain the chemistry behind carbon’s affinity for covalently bonding in organic compounds
- Provide examples of three types of carbohydrates, and identify the primary functions of carbohydrates in the body
- Discuss four types of lipids important in human functioning
- Describe the structure of proteins, and discuss their importance to human functioning
- Identify the building blocks of nucleic acids, and the roles of DNA, RNA, and ATP in human functioning

Organic compounds typically consist of groups of carbon atoms covalently bonded to hydrogen, usually oxygen, and often other elements as well. Created by living things, they are found throughout the world, in soils and seas, commercial products, and every cell of the human body. The four types most important to human structure and function are carbohydrates, lipids, proteins, and nucleotides. Before exploring these compounds, you need to first understand the chemistry of carbon.

The Chemistry of Carbon

What makes organic compounds ubiquitous is the chemistry of their carbon core. Recall that carbon atoms have four electrons in their valence shell, and that the octet rule dictates that atoms tend to react in such a way as to complete their valence shell with eight electrons. Carbon atoms do not complete their valence shells by donating or accepting four electrons. Instead, they readily share electrons via covalent bonds.

Commonly, carbon atoms share with other carbon atoms, often forming a long carbon chain referred to as a carbon skeleton. When they do share, however, they do not share all their electrons exclusively with each other. Rather, carbon atoms tend to share electrons with a variety of other elements, one of which is always hydrogen. Carbon and hydrogen groupings are called hydrocarbons. If you study the figures of organic compounds in the remainder of this chapter, you will see several with chains of hydrocarbons in one region of the compound.

Many combinations are possible to fill carbon’s four “vacancies.” Carbon may share electrons with oxygen or nitrogen or other atoms in a particular region of an organic compound. Moreover, the atoms to which carbon atoms bond may also be part of a functional group. A **functional group** is a group of atoms linked by strong covalent bonds and tending to function in chemical reactions as a single unit. You can think of functional groups as tightly knit “cliques” whose members are unlikely to be parted. Six functional groups are important in human physiology; these are the hydroxyl, carboxyl, amino, methyl, sulfhydryl, and phosphate groups ([link]).

Functional Groups Important in Human Physiology

Functional group	Structural formula	Importance
Hydroxyl	—O—H	Hydroxyl groups are polar. They are components of all four types of organic compounds discussed in this chapter. They are involved in dehydration synthesis and hydrolysis reactions.
Carboxyl	O—C—OH	Carboxyl groups act as acids and are found within fatty acids, amino acids, and many other acids.

Functional Groups Important in Human Physiology		
Functional group	**Structural formula**	**Importance**
Amino	—N—H_2	Amino groups act as bases and are found within amino acids, the building blocks of proteins.
Methyl	—C—H_3	Methyl groups in organic compounds tend to be stable and do not readily react, they are found within amino acids.
Sulfhydryl	—S—H	Sulfhydryl groups, called thiols, are polar and are found in the amino acid cysteine. They form disulfide bonds important in tertiary and quaternary protein structure. They are also important in the active site of some enzymes and are highly reactive with heavy metals.
Phosphate	—P—O_4^{2-}	Phosphate groups are polar and found within phospholipids and nucleotides.

Carbon’s affinity for covalent bonding means that many distinct and relatively stable organic molecules nevertheless readily form larger, more complex molecules. Any large molecule is referred to as **macromolecule** (macro- = “large”), and the organic compounds in this section all fit this description. However, some macromolecules are made up of several “copies” of single units called monomer (mono- = “one”; -mer = “part”). Like beads in a long necklace, these monomers link by covalent bonds to form long polymers (poly- = “many”). There are many examples of monomers and polymers among the organic compounds.

Monomers form polymers by engaging in dehydration synthesis (see [link]). As was noted earlier, this reaction results in the release of a molecule of water. Each monomer contributes: One gives up a hydrogen atom and the other gives up a hydroxyl group. Polymers are split into monomers by hydrolysis (-lysis = “rupture”). The bonds between their monomers are broken, via the donation of a molecule of water, which contributes a hydrogen atom to one monomer and a hydroxyl group to the other.

Isomers

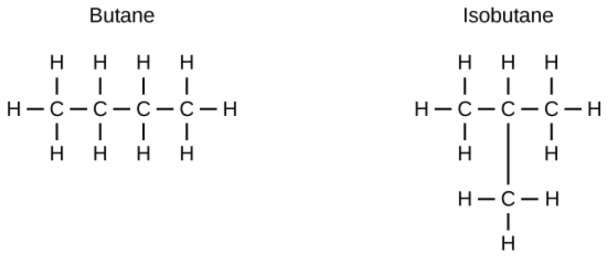
The three-dimensional placement of atoms and chemical bonds within organic molecules is central to understanding their chemistry. Molecules that share the same chemical formula but differ in the placement (structure) of their atoms and/or chemical bonds are known as **isomers**. **Structural isomers** (like butane and isobutene shown in [link]**a**) differ in the placement of their covalent bonds: both molecules have four carbons and ten hydrogens (C_4H_{10}), but the different arrangement of the atoms within the molecules leads to differences in their chemical properties. For example, due to their different chemical properties, butane is suited for use as a fuel for cigarette lighters and torches, whereas isobutene is suited for use as a refrigerant and a propellant in spray cans.

Geometric isomers, on the other hand, have similar placements of their covalent bonds but differ in how these bonds are made to the surrounding atoms, especially in carbon-to-carbon double bonds. In the simple molecule butene (C_4H_8), the two methyl groups (CH_3) can be on either side of the double covalent bond central to the molecule, as illustrated in [link]**b**. When the carbons are bound on the same side of the double bond, this is the *cis* configuration; if they are on opposite sides of the double bond, it is a *trans* configuration. In the *trans* configuration, the carbons form a more or less linear structure, whereas the carbons in the *cis* configuration make a bend (change in direction) of the carbon backbone.

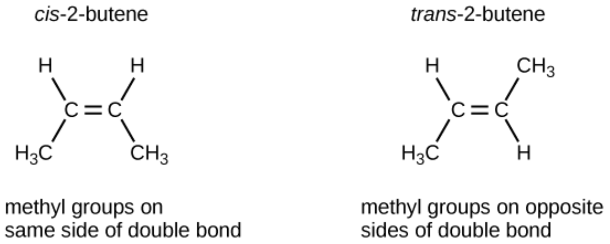
Enantiomers, (sometimes called stereoisomers) are mirror images of each other as illustrated in [link]**c**. They are like your left and right hands; they have the same structure but they cannot be superimposed. There are several naming systems for the two states. The most common in physiology is to use the prefix L for left-handed and D for right-handed enantiomers. Nineteen of the 20 common amino acids are designated as L-amino acids. This is important because an enantiomer often will react differently to another enantiomer substance depending on the compatibility of their configurations (L or D form). Because enzymes and receptors are so specific in their shape, only one form of the enantiomer is generally active in a physiological reaction. The other form may be inactive, weakly active, or may cause adverse reactions.

Art Connection:

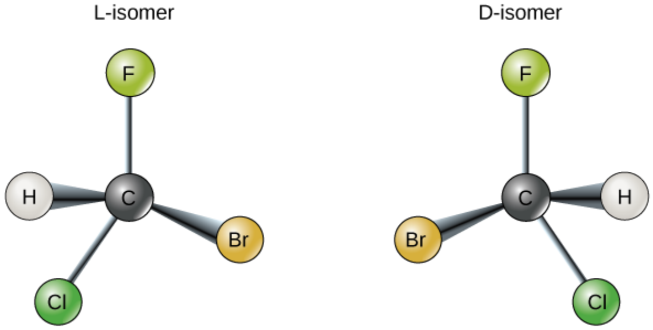
(a) Structural isomers



(b) Geometric isomers


cis-2-butene
trans-2-butene
H
C
H_3C
CH_3
methyl groups on same side of double bond
methyl groups on opposite sides of double bond

(c) Enantiomers



Molecules that have the same number and type of atoms arranged differently are called isomers. (a) Structural isomers have a different covalent arrangement of atoms. (b) Geometric isomers have a different arrangement of atoms around a double bond. (c) Enantiomers are mirror images of each other.

Carbohydrates

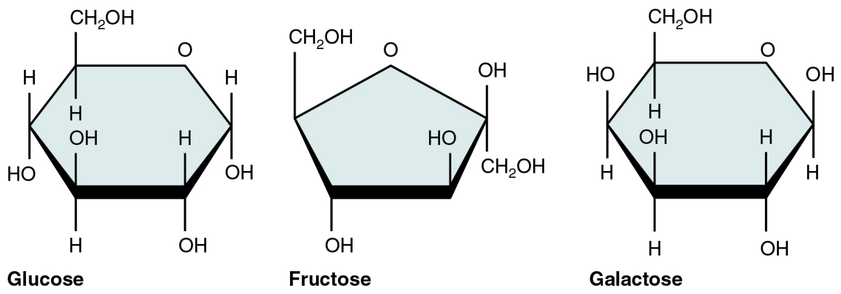
The term carbohydrate means “hydrated carbon.” Recall that the root hydro- indicates water. A **carbohydrate** is a molecule composed of carbon, hydrogen, and oxygen; in most carbohydrates, hydrogen and oxygen are found in the same two-to-one relative proportions they have in water. In fact, the chemical formula for a “generic” molecule of carbohydrate is $(CH_2O)_n$.

Carbohydrates are referred to as saccharides, a word meaning “sugars.” Three forms are important in the body. Monosaccharides are the monomers of carbohydrates. Disaccharides (di- = “two”) are made up of two monomers. **Polysaccharides** are the polymers, and can consist of hundreds to thousands of monomers.

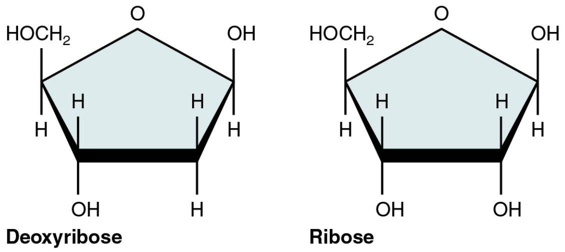
Monosaccharides

A **monosaccharide** is a monomer of carbohydrates. Five monosaccharides are important in the body. Three of these are the hexose sugars, so called because they each contain six atoms of carbon. These are glucose, fructose, and galactose, shown in [link]**a**. The remaining monosaccharides are the two pentose sugars, each of which contains five atoms of carbon. They are ribose and deoxyribose, shown in [link]**b**.

Five Important Monosaccharides



(a) Hexoses



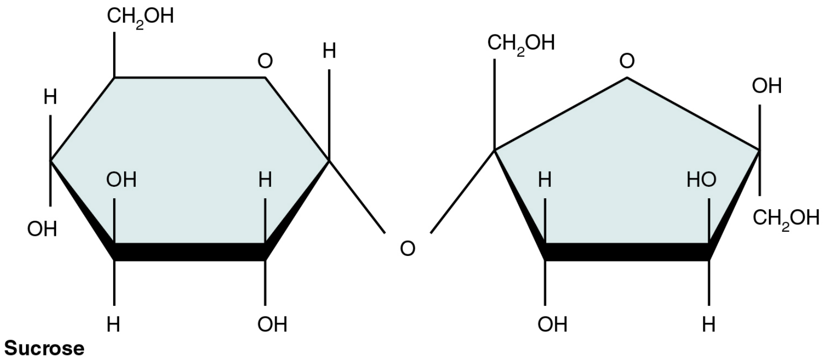
(b) Pentoses

Disaccharides

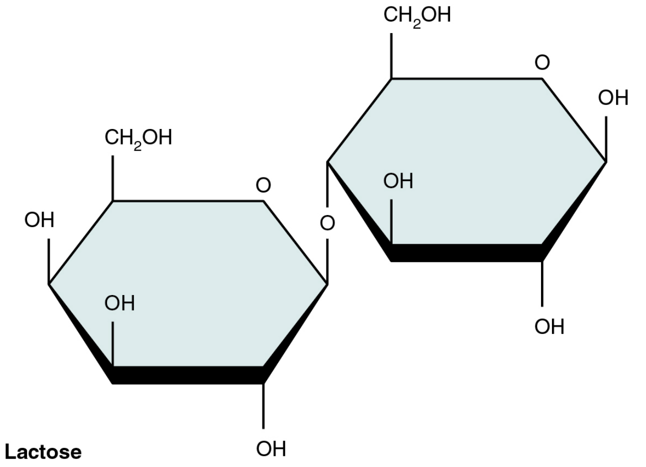
A **disaccharide** is a pair of monosaccharides. Disaccharides are formed via dehydration synthesis, and the bond linking them is referred to as a glycosidic bond (glyco- = “sugar”). Three disaccharides (shown in [link]) are important to humans. These are sucrose, commonly referred to as table sugar; lactose, or milk sugar; and maltose, or malt sugar. As you can tell from their common names, you consume these in your diet; however, your body cannot use them directly. Instead, in the digestive tract, they are split into their component monosaccharides via hydrolysis.

Three Important Disaccharides

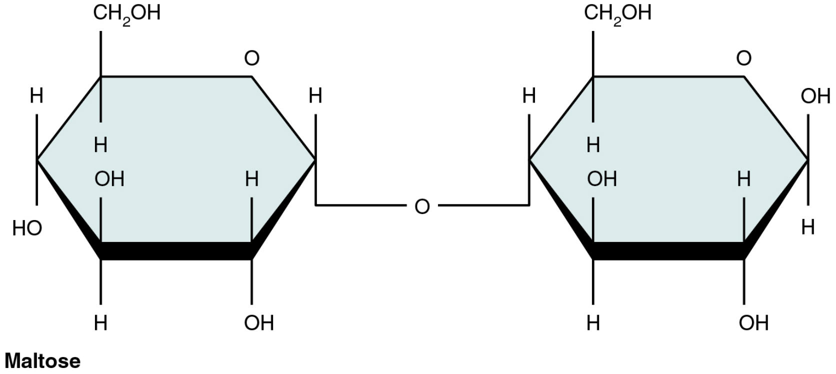
(a) The monosaccharides glucose and fructose bond to form sucrose



(b) The monosaccharides galactose and glucose bond to form lactose.



(c) Two glucose monosaccharides bond to form maltose.



All three important disaccharides form by dehydration synthesis.

Note:

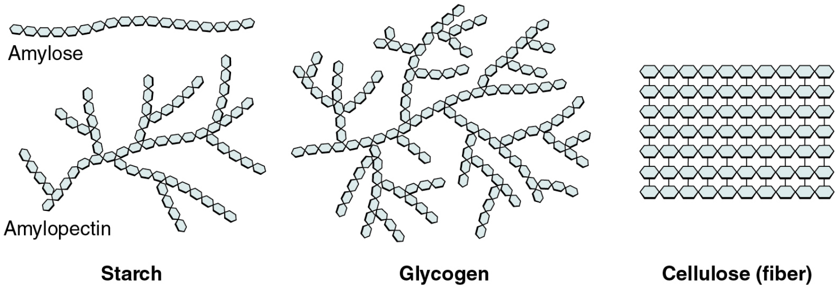
Watch this video to observe the formation of a disaccharide. What happens when water encounters a glycosidic bond?

Polysaccharides

Polysaccharides can contain a few to a thousand or more monosaccharides. Three are important to the body ([link]):

- Starches are polymers of glucose. They occur in long chains called amylose or branched chains called amylopectin, both of which are stored in plant-based foods and are relatively easy to digest.
- Glycogen is also a polymer of glucose, but it is stored in the tissues of animals, especially in the muscles and liver. It is not considered a dietary carbohydrate because very little glycogen remains in animal tissues after slaughter; however, the human body stores excess glucose as glycogen, again, in the muscles and liver.
- Cellulose, a polysaccharide that is the primary component of the cell wall of green plants, is the component of plant food referred to as “fiber”. In humans, cellulose/fiber is not digestible; however, dietary fiber has many health benefits. It helps you feel full so you eat less, it promotes a healthy digestive tract, and a diet high in fiber is thought to reduce the risk of heart disease and possibly some forms of cancer.
- Chitin (not shown) is a polysaccharide similar in structure to cellulose that makes up the hard exoskeleton of insects and crustaceans. Like cellulose it is not digestible by humans.

Three Important Polysaccharides



Three important polysaccharides are starches, glycogen, and fiber.

Functions of Carbohydrates

The body obtains carbohydrates from plant-based foods. Grains, fruits, and legumes and other vegetables provide most of the carbohydrate in the human diet, although lactose is found in dairy products.

Although most body cells can break down other organic compounds for fuel, all body cells can use glucose. Moreover, nerve cells (neurons) in the brain, spinal cord, and through the peripheral nervous system, as well as red blood cells, can use only glucose for fuel. In the breakdown of glucose for energy, molecules of adenosine triphosphate, better known as ATP, are produced. **Adenosine triphosphate (ATP)** is composed of a ribose sugar, an adenine base, and three phosphate groups. ATP releases free energy when its phosphate bonds are broken, and thus supplies ready energy to the cell. More ATP is produced in the presence of oxygen (O_2) than in pathways that do not use oxygen. The overall reaction for the conversion of the energy in glucose to energy stored in ATP can be written:

Equation:

$$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O + ATP$$

In addition to being a critical fuel source, carbohydrates are present in very small amounts in cells’ structure. For instance, some carbohydrate molecules bind with proteins to produce glycoproteins, and others combine with lipids to produce glycolipids, both of which are found in the membrane that encloses the contents of body cells.

Lipids

A **lipid** is one of a highly diverse group of compounds made up mostly of hydrocarbons. The few oxygen atoms they contain are often at the periphery of the molecule. Their nonpolar hydrocarbons make all lipids hydrophobic. In water, lipids do not form a true solution, but they may form an emulsion, which is the term for a mixture of solutions that do not mix well.

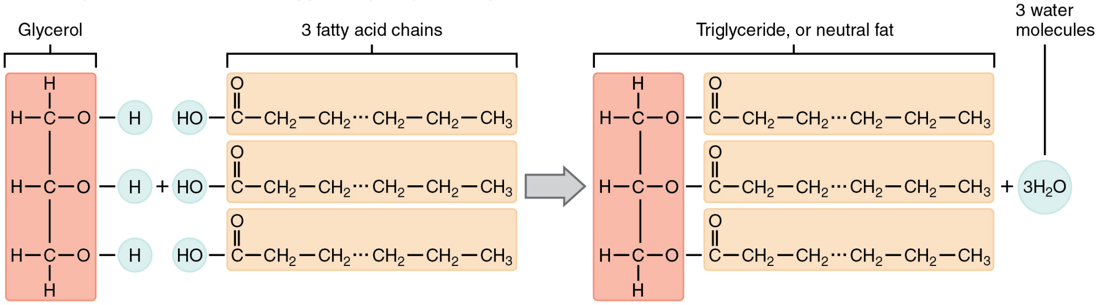
Triglycerides

A **triglyceride** is one of the most common dietary lipid groups, and the type found most abundantly in body tissues. This compound, which is commonly referred to as a fat, is formed from the synthesis of two types of molecules ([link]):

- A glycerol backbone at the core of triglycerides, consists of three carbon atoms.
- Three fatty acids, long chains of hydrocarbons with a carboxyl group and a methyl group at opposite ends, extend from each of the carbons of the glycerol.

Triglycerides

Three fatty acid chains are bound to glycerol by dehydration synthesis.

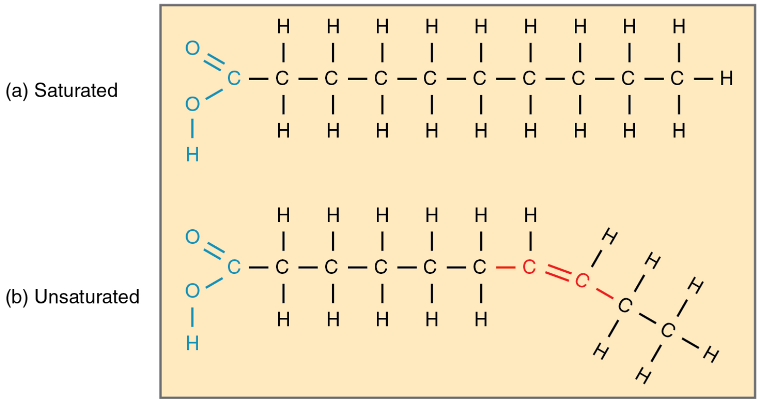


Triglycerides are composed of glycerol attached to three fatty acids via dehydration synthesis. Notice that glycerol gives up a hydrogen atom, and the carboxyl groups on the fatty acids each give up a hydroxyl group.

Triglycerides form via dehydration synthesis. Glycerol gives up hydrogen atoms from its hydroxyl groups at each bond, and the carboxyl group on each fatty acid chain gives up a hydroxyl group. A total of three water molecules are thereby released.

Fatty acid chains that have no double carbon bonds anywhere along their length and therefore contain the maximum number of hydrogen atoms are called saturated fatty acids. These straight, rigid chains pack tightly together and are solid or semi-solid at room temperature ([link]**a**). Butter and lard are examples, as is the fat found on a steak or in your own body. In contrast, fatty acids with one double carbon bond are kinked at that bond ([link]**b**). These monounsaturated fatty acids are therefore unable to pack together tightly, and are liquid at room temperature. Polyunsaturated fatty acids contain two or more double carbon bonds, and are also liquid at room temperature. Plant oils such as olive oil typically contain both mono- and polyunsaturated fatty acids.

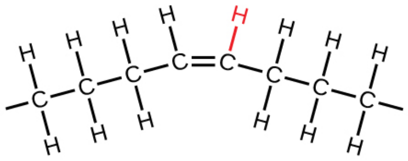
Fatty Acid Shapes



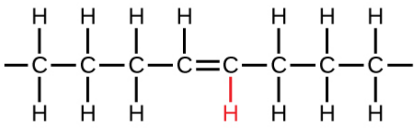
The level of saturation of a fatty acid affects its shape. (a) Saturated fatty acid chains are straight. (b) Unsaturated fatty acid chains are kinked.

Whereas a diet high in saturated fatty acids increases the risk of heart disease, a diet high in unsaturated fatty acids is thought to reduce the risk. This is especially true for the omega-3 unsaturated fatty acids found in cold-water fish such as salmon. These fatty acids have their first double carbon bond at the third hydrocarbon from the methyl group (referred to as the omega end of the molecule).

Finally, *trans* fatty acids found in some processed foods, including some stick and tub margarines, are thought to be even more harmful to the heart and blood vessels than saturated fatty acids. *Trans* fats are created from unsaturated fatty acids (such as corn oil) when chemically treated to produce partially hydrogenated fats ([link]).



***cis*-fat molecule**



***trans*-fat molecule**

During the hydrogenation process, the orientation around the double bonds is changed, making a trans-fat from a cis-fat. This changes the chemical properties of the molecule.

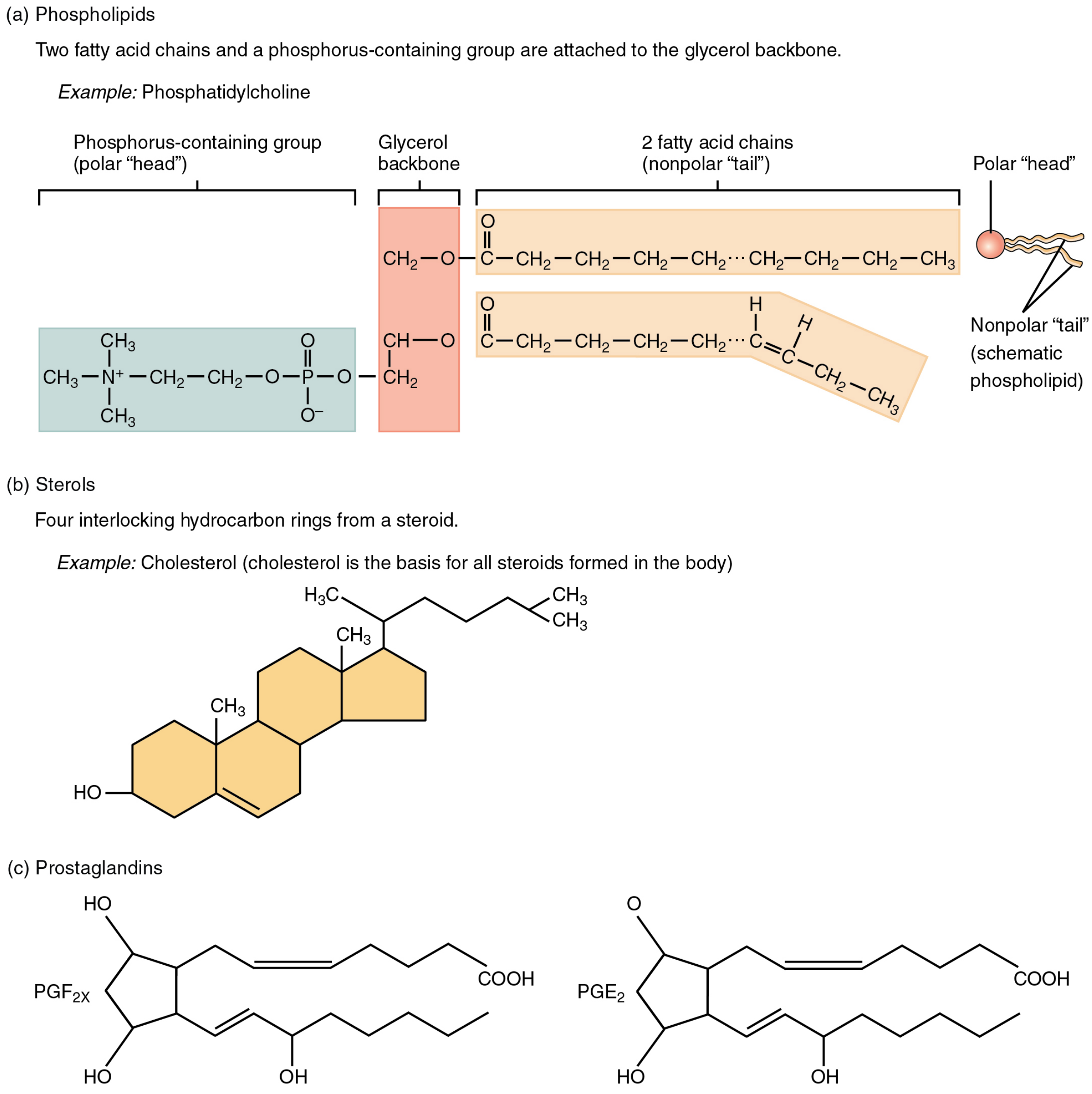
As a group, triglycerides are a major fuel source for the body. When you are resting or asleep, a majority of the energy used to keep you alive is derived from triglycerides stored in your fat (adipose) tissues. Triglycerides also fuel long, slow physical activity such as gardening or hiking, and contribute a modest percentage of energy for vigorous physical activity. Dietary fat also assists the absorption and transport of the nonpolar fat-soluble vitamins A, D, E, and K. Additionally, stored body fat protects and cushions the body's bones and internal organs, and acts as insulation to retain body heat.

Fatty acids are also components of glycolipids, which are sugar-fat compounds found in the cell membrane. Lipoproteins are compounds in which the hydrophobic triglycerides are packaged in protein envelopes for transport in body fluids.

Phospholipids

As its name suggests, a **phospholipid** is a bond between the glycerol component of a lipid and a phosphorous molecule. In fact, phospholipids are similar in structure to triglycerides. However, instead of having three fatty acids, a phospholipid is generated from a diglyceride, a glycerol with just two fatty acid chains ([link]). The third binding site on the glycerol is taken up by the phosphate group, which in turn is attached to a polar "head" region of the molecule. Recall that triglycerides are nonpolar and hydrophobic. This still holds for the fatty acid portion of a phospholipid compound. However, the phosphate-containing group at the head of the compound is polar and thereby hydrophilic. In other words, one end of the molecule can interact with oil, and the other end with water. This makes phospholipids ideal emulsifiers, compounds that help disperse fats in aqueous liquids, and enables them to interact with both the watery interior of cells and the watery solution outside of cells as components of the cell membrane.

Other Important Lipids


(a) Phospholipids
Two fatty acid chains and a phosphorus-containing group are attached to the glycerol backbone.
Example: Phosphatidylcholine
Phosphorus-containing group (polar "head")
Glycerol backbone
2 fatty acid chains (nonpolar "tail")
Polar "head"
Nonpolar "tail" (schematic phospholipid)
CH3
CH3—N+—CH2—CH2—O—P—O
CH3
O
O−
CH2—O
CH—O
CH2
O
C—CH2—CH2—CH2—CH2··· CH2—CH2—CH2—CH3
O
C—CH2—CH2—CH2—CH2··· C=C—CH2—CH3
H
H
(b) Sterols
Four interlocking hydrocarbon rings from a steroid.
Example: Cholesterol (cholesterol is the basis for all steroids formed in the body)
H3C
CH3
CH3
CH3
CH3
HO
(c) Prostaglandins
HO
PGF2X
HO
OH
COOH
O
PGE2
HO
OH
COOH

(a) Phospholipids are composed of two fatty acids, glycerol, and a phosphate group. (b) Sterols are ring-shaped lipids. Shown here is cholesterol. (c) Prostaglandins are derived from unsaturated fatty acids. Prostaglandin E2 (PGE2) includes hydroxyl and carboxyl groups.

Steroids

A **steroid** compound (referred to as a sterol) has as its foundation a set of four hydrocarbon rings bonded to a variety of other atoms and molecules (see [link]**b**). Although both plants and animals synthesize sterols, the type that makes the most important contribution to human structure and function is cholesterol, which is synthesized by the liver in humans and animals and is also present in most animal-based foods. Like other lipids, cholesterol’s hydrocarbons make it hydrophobic; however, it has a polar hydroxyl head that is hydrophilic. Cholesterol is an important component of bile acids, compounds that help emulsify dietary fats. In fact, the word root chole- refers to bile. Cholesterol is also a building block of many hormones, signaling molecules that the body releases to regulate processes at distant sites. Finally, like phospholipids, cholesterol molecules are found in the cell membrane, where their hydrophobic and hydrophilic regions help regulate the flow of substances into and out of the cell.

Prostaglandins

Like a hormone, a **prostaglandin** is one of a group of signaling molecules, but prostaglandins are derived from unsaturated fatty acids (see [link]**c**). One reason that the omega-3 fatty acids found in fish are beneficial is that they stimulate the production of certain prostaglandins that help regulate aspects of blood pressure and inflammation, and thereby reduce the risk for heart disease. Prostaglandins also sensitize nerves to pain. One class of pain-relieving medications called nonsteroidal anti-inflammatory drugs (NSAIDs) works by reducing the effects of prostaglandins.

Proteins

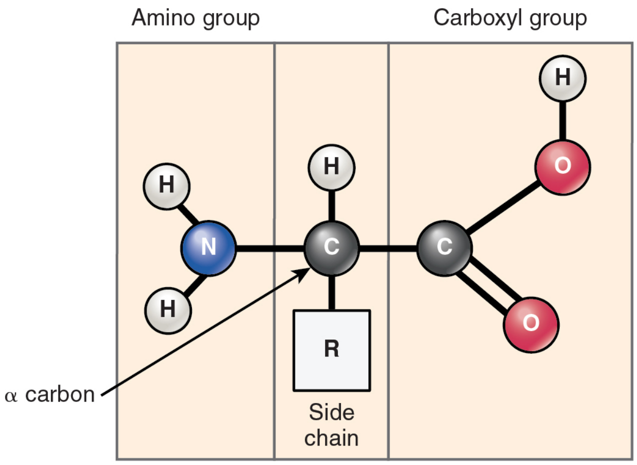
You might associate proteins with muscle tissue, but in fact, proteins are critical components of all tissues and organs. A **protein** is an organic molecule composed of amino acids linked by peptide bonds. Proteins include the keratin in the epidermis of skin that protects underlying tissues, the collagen found in the dermis of skin, in bones, and in the meninges that cover the brain and spinal cord. Proteins are also components of many of the body’s functional chemicals, including digestive enzymes in the digestive tract, antibodies, the neurotransmitters that neurons use to communicate with other cells, and the peptide-based hormones that regulate certain body functions (for instance, growth hormone). While carbohydrates and lipids are composed of hydrocarbons and oxygen, all proteins also contain nitrogen (N), and many contain sulfur (S), in addition to carbon, hydrogen, and oxygen.

Microstructure of Proteins

Proteins are polymers made up of nitrogen-containing monomers called amino acids. An **amino acid** is a molecule composed of an amino group and a carboxyl group, together with a variable side chain. Just 20 different amino acids contribute to nearly all of the thousands of different proteins important in human structure and function. Body proteins contain a unique combination of a few dozen to a few hundred of these 20 amino acid monomers. All 20 of these amino acids share a similar structure ([link]). All consist of a central carbon atom to which the following are bonded:

- a hydrogen atom
- an alkaline (basic) amino group NH_2 (see [link])
- an acidic carboxyl group COOH (see [link])
- a variable group

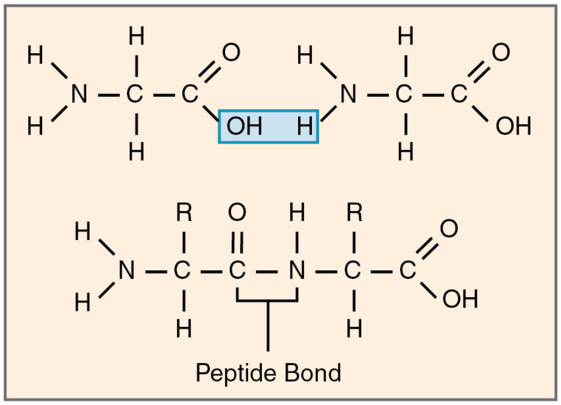
Structure of an Amino Acid



Notice that all amino acids contain both an acid (the carboxyl group) and a base (the amino group) (amine = “nitrogen-containing”). For this reason, they make excellent buffers, helping the body regulate acid–base balance. What distinguishes the 20 amino acids from one another is their variable group, which is referred to as a side chain or an R-group. This group can vary in size and can be polar or nonpolar, giving each amino acid its unique characteristics. For example, the side chains of two amino acids—cysteine and methionine—contain sulfur. Sulfur does not readily participate in hydrogen bonds, whereas all other amino acids do. This variation influences the way that proteins containing cysteine and methionine are assembled.

Amino acids join via dehydration synthesis to form protein polymers ([link]). The unique bond holding amino acids together is called a peptide bond. A **peptide bond** is a covalent bond between two amino acids that forms by dehydration synthesis. A peptide, in fact, is a very short chain of amino acids. Strands containing fewer than about 100 amino acids are generally referred to as polypeptides rather than proteins.

Peptide Bond



Different amino acids join together to form peptides, polypeptides, or proteins via dehydration synthesis. The bonds between the amino acids are peptide bonds.

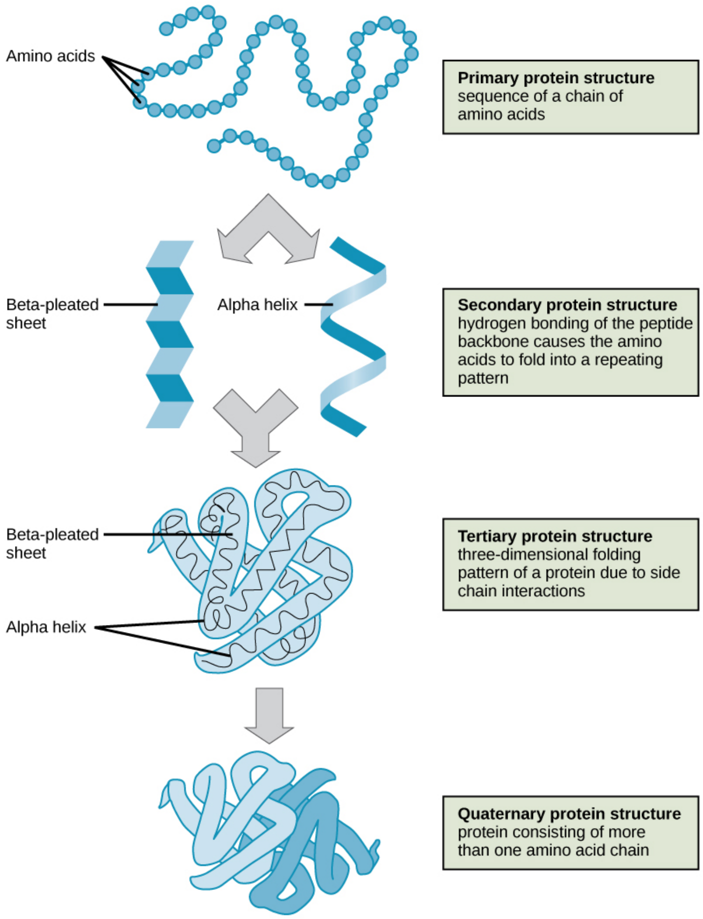
The body is able to synthesize most of the amino acids from components of other molecules; however, nine cannot be synthesized and have to be consumed in the diet. These are known as the essential amino acids.

Free amino acids available for protein construction are said to reside in the amino acid pool within cells. Structures within cells use these amino acids when assembling proteins. If a particular essential amino acid is not available in sufficient quantities in the amino acid pool, however, synthesis of proteins containing it can slow or even cease.

Shape of Proteins

Primary Structure

Just as a fork cannot be used to eat soup and a spoon cannot be used to spear meat, a protein's shape is essential to its function. A protein's shape is determined, most fundamentally, by the sequence of amino acids of which it is made ([link]). The sequence is called the primary structure of the protein.


Amino acids
Primary protein structure
sequence of a chain of amino acids
Beta-pleated sheet
Alpha helix
Secondary protein structure
hydrogen bonding of the peptide backbone causes the amino acids to fold into a repeating pattern
Beta-pleated sheet
Alpha helix
Tertiary protein structure
three-dimensional folding pattern of a protein due to side chain interactions
Quaternary protein structure
protein consisting of more than one amino acid chain

The four levels of protein structure can be observed in these illustrations. (credit: modification of work by National Human Genome Research Institute)

Secondary Structure

Although some polypeptides exist as linear chains, most are twisted or folded into more complex secondary structures that form when bonding occurs between amino acids with different properties at different regions of the polypeptide. The most common secondary structure is a spiral called an alpha-helix. If you were to take a length of string and simply twist it into a spiral, it would not hold the shape. Similarly, a strand of amino acids could not maintain a stable spiral shape without the help of hydrogen bonds, which create bridges between different regions of the same strand. Less commonly, a polypeptide chain can form a beta-pleated sheet, in which hydrogen bonds form bridges between different regions of a single polypeptide that has folded back upon itself, or between two or more adjacent polypeptide chains.

Tertiary Structure

The secondary structure of proteins further folds into a compact three-dimensional shape, referred to as the protein's tertiary structure. In this configuration, amino acids that had been very distant in the primary chain can be brought quite close via hydrogen bonds or, in proteins containing cysteine, via disulfide bonds. A **disulfide bond** is a covalent bond between sulfur atoms in a polypeptide.

Quaternary Structure

Often, two or more separate polypeptides bond to form an even larger protein with a quaternary structure. The polypeptide subunits forming a quaternary structure can be identical or different. For instance, hemoglobin, the protein found in red blood cells is composed of four tertiary polypeptides, two of which are called alpha chains and two of which are called beta chains.

When they are exposed to extreme heat, acids, bases, and certain other substances, proteins will denature. **Denaturation** is a change in the structure of a molecule through physical or chemical means. Denatured proteins lose their functional shape and are no longer able to carry out their jobs. An everyday example of protein denaturation is the curdling of milk when acidic lemon juice is added.

The contribution of the shape of a protein to its function can hardly be exaggerated. For example, the long, slender shape of protein strands that make up muscle tissue is essential to their ability to contract (shorten) and relax (lengthen). As another example, bones contain long threads of a protein called collagen that acts as scaffolding upon which bone minerals are deposited. These elongated proteins, called fibrous proteins, are strong and durable and typically hydrophobic.

In contrast, globular proteins are globes or spheres that tend to be highly reactive and are hydrophilic. The hemoglobin proteins packed into red blood cells are an example; however, globular proteins are abundant throughout the body, playing critical roles in most body functions. Enzymes, introduced earlier as protein catalysts, are examples of this.

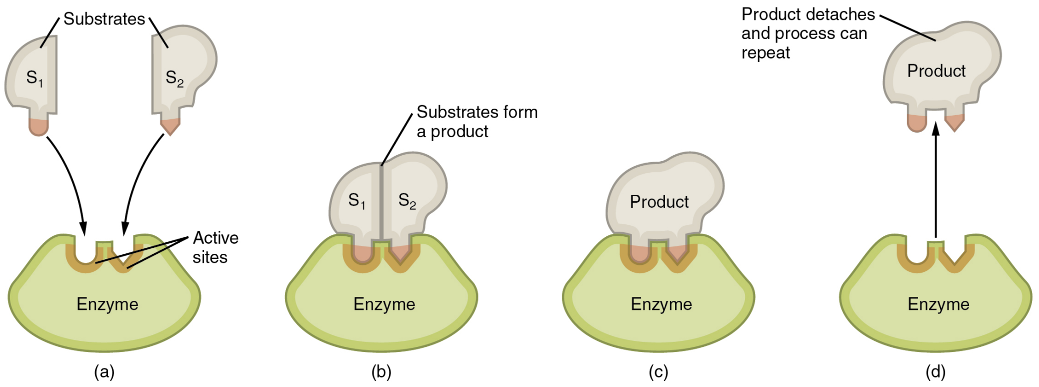
Living organisms are made of proteins and the products of protein activities. The infinite array of complex three dimensional shapes and electrical charge patterns proteins can take provide the basis for both cellular and organismic structure, movement, regulation, facilitation, communication, transport, defense, and storage. The next section takes a closer look at how enzymes facilitate cell activities.

Proteins Function as Enzymes

If you were trying to type a paper, and every time you hit a key on your laptop there was a delay of six or seven minutes before you got a response, you would probably get a new laptop. In a similar way, without enzymes to catalyze chemical reactions, the human body would be nonfunctional. It functions only because enzymes function.

Enzymatic reactions—chemical reactions catalyzed by enzymes—begin when substrates bind to the enzyme. A **substrate** is a reactant in an enzymatic reaction. This occurs on regions of the enzyme known as active sites ([link]). Any given enzyme catalyzes just one type of chemical reaction. This characteristic, called specificity, is due to the fact that a substrate with a particular shape and electrical charge can bind only to an active site corresponding to that substrate.

Steps in an Enzymatic Reaction



(a) Substrates approach active sites on enzyme. (b) Substrates bind to active sites, producing an enzyme–substrate complex. (c) Changes internal to the enzyme–substrate complex facilitate interaction of the substrates. (d) Products are released and the enzyme returns to its original form, ready to facilitate another enzymatic reaction.

Binding of a substrate produces an enzyme–substrate complex. It is likely that enzymes speed up chemical reactions in part because the enzyme–substrate complex undergoes a set of temporary and reversible changes that cause the substrates to be oriented toward each other in an optimal position to facilitate their interaction. This promotes increased reaction speed. The enzyme then releases the product(s), and resumes its original shape. The enzyme is then free to engage in the process again, and will do so as long as substrate remains.

Other Functions of Proteins

As was noted earlier, the basic and acidic components enable proteins to function as buffers in maintaining acid–base balance, but they also help regulate fluid–electrolyte balance. Proteins attract fluid, and a healthy concentration of proteins in the blood, the cells, and the spaces between cells helps ensure a balance of fluids in these various “compartments.” Moreover, proteins in the cell membrane help to transport electrolytes in and out of the cell, keeping these ions in a healthy balance. Like lipids, proteins can bind with carbohydrates. They can thereby produce glycoproteins or proteoglycans, both of which have many functions in the body.

The body can use proteins for energy when carbohydrate and fat intake is inadequate, and stores of glycogen and adipose tissue become depleted. However, since there is no storage site for protein except functional tissues, using protein for energy causes tissue breakdown, and results in body wasting.

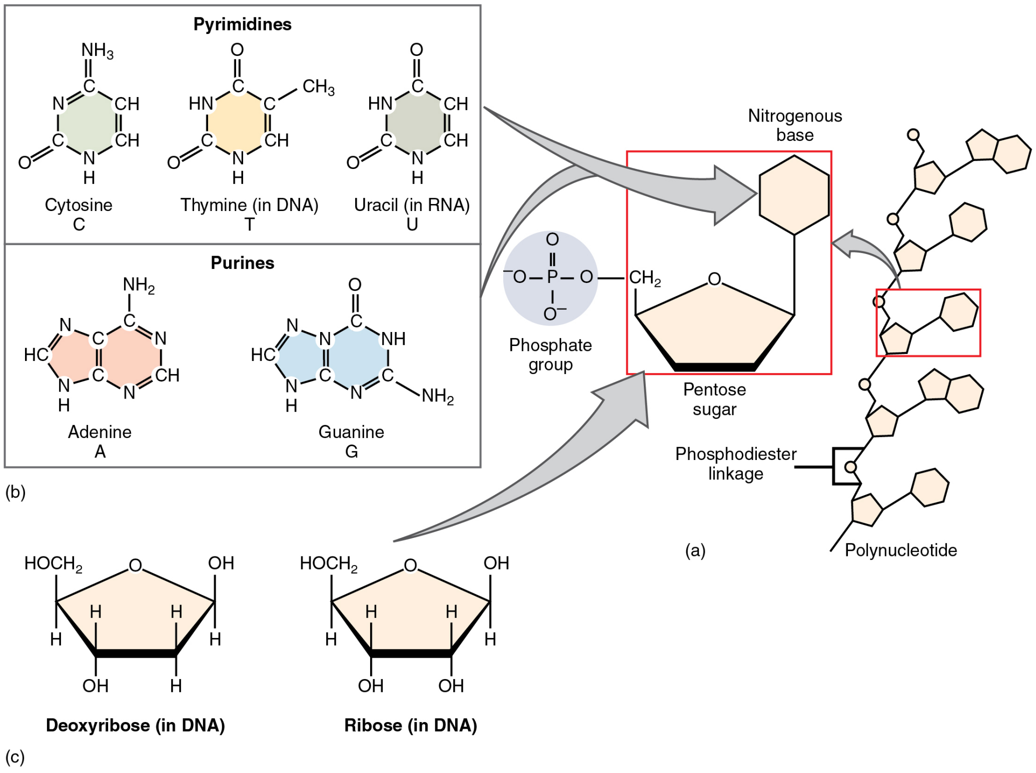
Nucleotides

The fourth type of organic compound important to human structure and function are the nucleotides ([link]). A **nucleotide** is one of a class of organic compounds composed of three subunits:

- one or more phosphate groups
- a pentose sugar: either deoxyribose or ribose
- a nitrogen-containing base: adenine, cytosine, guanine, thymine, or uracil

Nucleotides can be assembled into nucleic acids (DNA or RNA) or the energy compound adenosine triphosphate.

Nucleotides



(a) The building blocks of all nucleotides are one or more phosphate groups, a pentose sugar, and a nitrogen-containing base. (b) The nitrogen-containing bases of nucleotides. (c) The two pentose sugars of DNA and RNA.

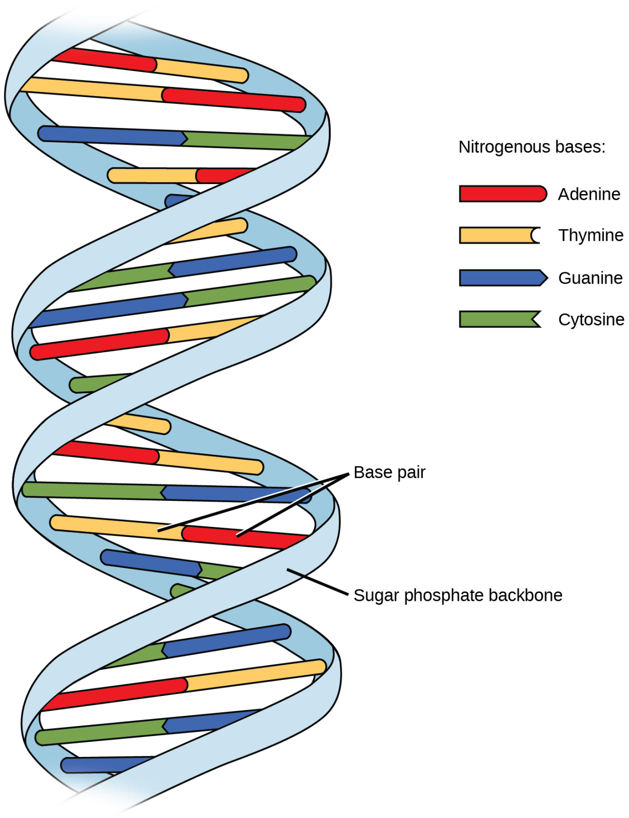
Nucleic Acids

The nucleic acids differ in their type of pentose sugar. **Deoxyribonucleic acid (DNA)** is nucleotide that stores genetic information. DNA contains deoxyribose (so-called because it has one less atom of oxygen than ribose) plus one phosphate group and one nitrogen-containing base. The “choices” of base for DNA are adenine, cytosine, guanine, and thymine. **Ribonucleic acid (RNA)** is a ribose-containing nucleotide that helps manifest the genetic code as protein. RNA contains ribose, one phosphate group, and one nitrogen-containing base, but the “choices” of base for RNA are adenine, cytosine, guanine, and uracil.

The nitrogen-containing bases adenine and guanine are classified as purines. A **purine** is a nitrogen-containing molecule with a double ring structure, which accommodates several nitrogen atoms. The bases cytosine, thymine (found in DNA only) and uracil (found in RNA only) are pyramidines. A **pyramidine** is a nitrogen-containing base with a single ring structure

Bonds formed by dehydration synthesis between the pentose sugar of one nucleic acid monomer and the phosphate group of another form a “backbone,” from which the components’ nitrogen-containing bases protrude. In DNA, two such backbones attach at their protruding bases via hydrogen bonds. These twist to form a shape known as a double helix ([link]). The sequence of nitrogen-containing bases within a strand of DNA form the genes that act as a molecular code instructing cells in the assembly of amino acids into proteins. Humans have almost 22,000 genes in their DNA, locked up in the 46 chromosomes inside the nucleus of each cell (except red blood cells which lose their nuclei during development). These genes carry the genetic code to build one’s body, and are unique for each individual except identical twins.

DNA



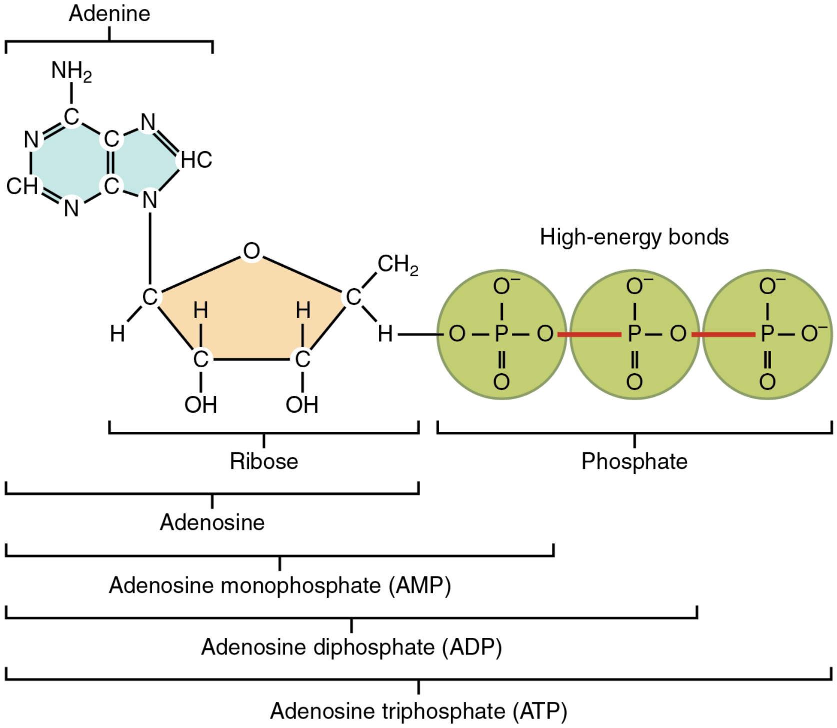
In the DNA double helix, two strands attach via hydrogen bonds between the bases of the component nucleotides.

In contrast, RNA consists of a single strand of sugar-phosphate backbone studded with bases. Messenger RNA (mRNA) is created during protein synthesis to carry the genetic instructions from the DNA to the cell’s protein manufacturing plants in the cytoplasm, the ribosomes.

Adenosine Triphosphate

The nucleotide adenosine triphosphate (ATP), is composed of a ribose sugar, an adenine base, and three phosphate groups ([link]). ATP is classified as a high energy compound because the two covalent bonds linking its three phosphates store a significant amount of potential energy. In the body, the energy released from these high energy bonds helps fuel the body's activities, from muscle contraction to the transport of substances in and out of cells to anabolic chemical reactions.

Structure of Adenosine Triphosphate (ATP)



When a phosphate group is cleaved from ATP, the products are adenosine diphosphate (ADP) and inorganic phosphate (P_i). This hydrolysis reaction can be written:

Equation:

$$\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_\text{i} + \text{energy}$$

Removal of a second phosphate leaves adenosine monophosphate (AMP) and two phosphate groups. Again, these reactions also liberate the energy that had been stored in the phosphate-phosphate bonds. They are reversible, too, as when ADP undergoes phosphorylation. **Phosphorylation** is the addition of a phosphate group to an organic compound, in this case, resulting in ATP. In such cases, the same level of energy that had been released during hydrolysis must be reinvested to power dehydration synthesis.

Cells can also transfer a phosphate group from ATP to another organic compound. For example, when glucose first enters a cell, a phosphate group is transferred from ATP, forming glucose phosphate ($C_6H_{12}O_6$—P) and ADP. Once glucose is phosphorylated in this way, it can be stored as glycogen or metabolized for immediate energy.

Chapter Review

Organic compounds essential to human functioning include carbohydrates, lipids, proteins, and nucleotides. These compounds are said to be organic because they contain both carbon and hydrogen. Carbon atoms in organic compounds readily share electrons with hydrogen and other atoms, usually oxygen, and sometimes nitrogen. Carbon atoms also may bond with one or more functional groups such as carboxyls, hydroxyls, aminos, or phosphates. Monomers are single units of organic compounds. They bond by dehydration synthesis to form polymers, which can in turn be broken by hydrolysis.

Carbohydrate compounds provide essential body fuel. Their structural forms include monosaccharides such as glucose, disaccharides such as lactose, and polysaccharides, including starches (polymers of glucose), glycogen (the storage form of glucose), and fiber. All body cells can use glucose for fuel. It is converted via an oxidation-reduction reaction to ATP.

Lipids are hydrophobic compounds that provide body fuel and are important components of many biological compounds. Triglycerides are the most abundant lipid in the body, and are composed of a glycerol backbone attached to three fatty acid chains. Phospholipids are compounds composed of a diglyceride with a phosphate group attached at the molecule’s head. The result is a molecule with polar and nonpolar regions. Steroids are lipids formed of four hydrocarbon rings. The most important is cholesterol. Prostaglandins are signaling molecules derived from unsaturated fatty acids.

Proteins are critical components of all body tissues. They are made up of monomers called amino acids, which contain nitrogen, joined by peptide bonds. Protein shape is critical to its function. Most body proteins are globular. An example is enzymes, which catalyze chemical reactions.

Nucleotides are compounds with three building blocks: one or more phosphate groups, a pentose sugar, and a nitrogen-containing base. DNA and RNA are nucleic acids that function in protein synthesis. ATP is the body's fundamental molecule of energy transfer. Removal or addition of phosphates releases or invests energy.

Interactive Link Questions

Exercise:

Problem:

Watch this video to observe the formation of a disaccharide. What happens when water encounters a glycosidic bond?

Solution:

The water hydrolyses, or breaks, the glycosidic bond, forming two monosaccharides.

Review Questions

Exercise:

Problem:$C_6H_{12}O_6$ is the chemical formula for a ________.

a. polymer of carbohydrate
b. pentose monosaccharide
c. hexose monosaccharide
d. all of the above

Solution:

C

Exercise:

Problem:

What organic compound do brain cells primarily rely on for fuel?

a. glucose
b. glycogen
c. galactose
d. glycerol

Solution:

A

Exercise:

Problem:

Which of the following is a functional group that is part of a building block of proteins?

a. phosphate
b. adenine
c. amino
d. ribose

Solution:

C

Exercise:

Problem:

A pentose sugar is a part of the monomer used to build which type of macromolecule?

a. polysaccharides
b. nucleic acids

c. phosphorylated glucose
d. glycogen

Solution:

B

Exercise:

Problem:A phospholipid ________.

a. has both polar and nonpolar regions
b. is made up of a triglyceride bonded to a phosphate group
c. is a building block of ATP
d. can donate both cations and anions in solution

Solution:

A

Exercise:

Problem:

In DNA, nucleotide bonding forms a compound with a characteristic shape known as a(n) ________.

a. beta chain
b. pleated sheet
c. alpha helix
d. double helix

Solution:

D

Exercise:

Problem: Uracil ________.

a. contains nitrogen
b. is a pyrimidine
c. is found in RNA
d. all of the above

Solution:

D

Exercise:

Problem:

The ability of an enzyme’s active sites to bind only substrates of compatible shape and charge is known as ________.

a. selectivity
b. specificity
c. subjectivity
d. specialty

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

If the disaccharide maltose is formed from two glucose monosaccharides, which are hexose sugars, how many atoms of carbon, hydrogen, and oxygen does maltose contain and why?

Solution:

Maltose contains 12 atoms of carbon, but only 22 atoms of hydrogen and 11 atoms of oxygen, because a molecule of water is removed during its formation via dehydration synthesis.

Exercise:

Problem:

Once dietary fats are digested and absorbed, why can they not be released directly into the bloodstream?

Solution:

All lipids are hydrophobic and unable to dissolve in the watery environment of blood. They are packaged into lipoproteins, whose outer protein envelope enables them to transport fats in the bloodstream.

Glossary

adenosine triphosphate (ATP)	nucleotide containing ribose and an adenine base that is essential in energy transfer
amino acid	building block of proteins; characterized by an amino and carboxyl functional groups and a variable side-chain
carbohydrate	class of organic compounds built from sugars, molecules containing carbon, hydrogen, and oxygen in a 1-2-1 ratio
denaturation	change in the structure of a molecule through physical or chemical means

deoxyribonucleic acid (DNA)

deoxyribose-containing nucleotide that stores genetic information

disaccharide

pair of carbohydrate monomers bonded by dehydration synthesis via a glycosidic bond

disulfide bond

covalent bond formed within a polypeptide between sulfide groups of sulfur-containing amino acids, for example, cysteine

functional group

group of atoms linked by strong covalent bonds that tends to behave as a distinct unit in chemical reactions with other atoms

lipid

class of nonpolar organic compounds built from hydrocarbons and distinguished by the fact that they are not soluble in water

macromolecule

large molecule formed by covalent bonding

monosaccharide

monomer of carbohydrate; also known as a simple sugar

nucleotide

class of organic compounds composed of one or more phosphate groups, a pentose sugar, and a base

peptide bond

covalent bond formed by dehydration synthesis between two amino acids

phospholipid

a lipid compound in which a phosphate group is combined with a diglyceride

phosphorylation

addition of one or more phosphate groups to an organic compound

polysaccharide
compound consisting of more than two carbohydrate monomers bonded by dehydration synthesis via glycosidic bonds

prostaglandin
lipid compound derived from fatty acid chains and important in regulating several body processes

protein
class of organic compounds that are composed of many amino acids linked together by peptide bonds

purine
nitrogen-containing base with a double ring structure; adenine and guanine

pyrimidine
nitrogen-containing base with a single ring structure; cytosine, thiamine, and uracil

ribonucleic acid (RNA)
ribose-containing nucleotide that helps manifest the genetic code as protein

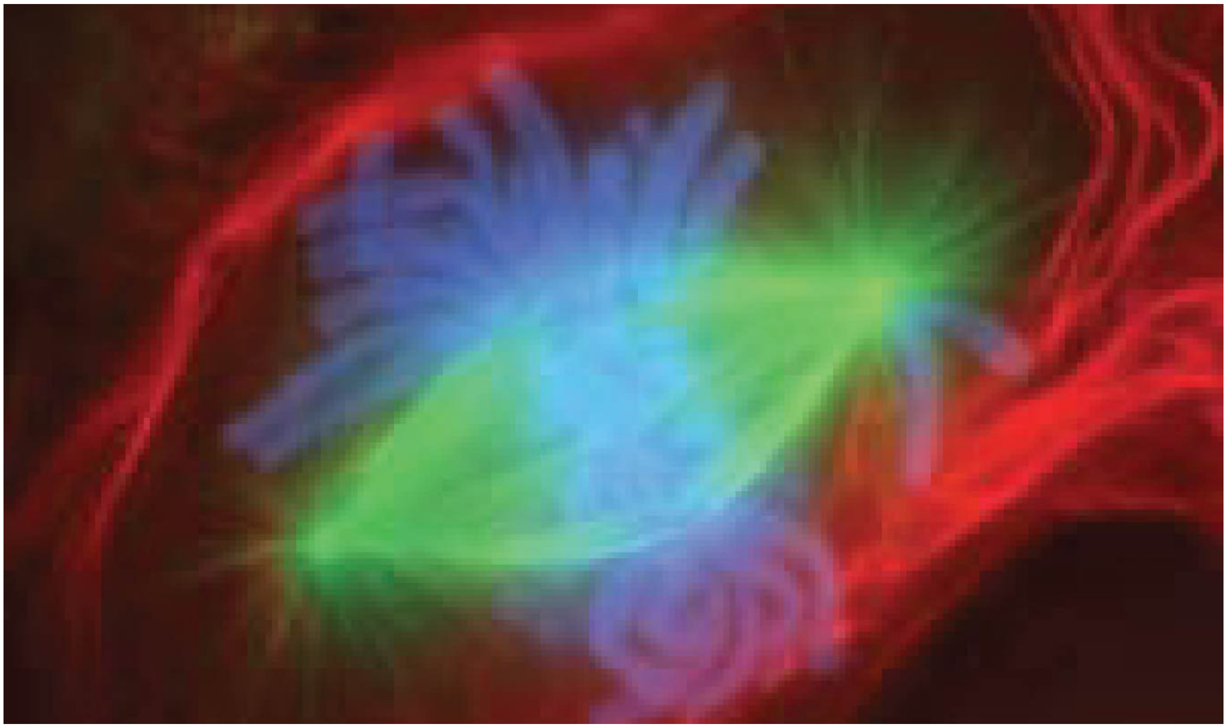
steroid
(also, sterol) lipid compound composed of four hydrocarbon rings bonded to a variety of other atoms and molecules

substrate
reactant in an enzymatic reaction

triglyceride
lipid compound composed of a glycerol molecule bonded with three fatty acid chains

Introduction to Cells
class="introduction"
Flourescence-stained Cell Undergoing Mitosis

A lung cell from a newt, commonly studied for its similarity to human lung cells, is stained with fluorescent dyes. The green stain reveals mitotic spindles, red is the cell membrane and part of the cytoplasm, and the structures that appear light blue are chromosomes. This cell is in anaphase of mitosis. (credit: “Mortadelo2005”/Wikimedia Commons)



Note:
Chapter Objectives
After studying this chapter, you will be able to:

- Describe the structure and function of the cell membrane, including its regulation of materials into and out of the cell
- Describe the functions of the various cytoplasmic organelles
- Explain the structure and contents of the nucleus, as well as the process of DNA replication
- Explain the process by which a cell builds proteins using the DNA code
- List the stages of the cell cycle in order, including the steps of cell division in both somatic cells
- Discuss how a cell differentiates and becomes more specialized
- List the morphological and physiological characteristics of some representative cell types in the human body

You developed from a single fertilized egg cell into the complex organism containing trillions of cells that you see when you look in a mirror. During this developmental process, early, undifferentiated cells differentiate and become specialized in their structure and function. These different cell types form specialized tissues that work in concert to perform all of the functions necessary for the living organism. Cellular and developmental biologists study how the continued division of a single cell leads to such complexity and differentiation.

Consider the difference between a structural cell in the skin and a nerve cell. A structural skin cell may be shaped like a flat plate (squamous) and live only for a short time before it is shed and replaced. Packed tightly into rows and sheets, the squamous skin cells provide a protective barrier for the cells and tissues that lie beneath. A nerve cell, on the other hand, may be shaped something like a star, sending out long processes up to a meter in length and may live for the entire lifetime of the organism. With their long winding appendages, nerve cells can communicate with one another and with other types of body cells and send rapid signals that inform the organism about its environment and allow it to interact with that environment. These differences illustrate one very important theme that is consistent at all organizational levels of biology: the form of a structure is optimally suited to perform particular functions assigned to that structure. Keep this theme in mind as you tour the inside of a cell and are introduced to the various types of cells in the body.

A primary responsibility of each cell is to maintain its own homeostasis and thereby contribute to the homeostasis of the overall organism. Homeostasis is a term used in biology that refers to a dynamic state of balance within parameters that are compatible with life. For example, living cells require a water-based environment to survive in, and there are various physical (anatomical) and physiological mechanisms that keep all of the trillions of living cells in the human body moist. This is one aspect of homeostasis. When a particular parameter, such as blood pressure or blood oxygen content, moves far enough *out* of homeostasis (generally becoming too high or too low), illness or disease—and sometimes death—inevitably results.

The concept of a cell started with microscopic observations of dead cork tissue by scientist Robert Hooke in 1665. Without realizing their function or importance, Hook coined the term “cell” based on the resemblance of the small subdivisions in the cork to the rooms that monks inhabited, called cells. About ten years later, Antonie van Leeuwenhoek became the first person to observe living and moving cells under a microscope. In the century that followed, the theory that cells represented the basic unit of life would develop. These tiny fluid-filled sacs house components responsible for the thousands of biochemical reactions necessary for an organism to grow and survive. In this chapter, you will learn about the major components and functions of a prototypical, generalized cell and discover some of the different types of cells in the human body.

The common microanatomy of eukaryotic cells (the type humans have) can be divided into three major structural-functional units based on what is visible with a light microscope: the cell membrane, the cytoplasm, and the cell nucleus. The cell membrane creates the boundary between the inside and outside of the cell. The nucleus, surrounded by the nuclear membrane, separates the genetic material (DNA) from the rest of the interior of the cell.

Cytoplasm refers to the interior cellular contents not bounded by the nuclear membrane. Other cellular structures are visible with a light microscope, but their presence varies depending on the specific type of cell.

Electron microscopes have allowed cellular anatomists to see much smaller structures. It now is clear that cytoplasm in not a uniform mass but contains many smaller components. Just as our body has specialized organs to perform complex functions, so cells have compartmentalized structures called organelles (little organs) that perform the specialized functions of the cell. So cytoplasm consists of two major components: the solid organelles and the liquid cytosol where the chemical reactions take place. Everything our body does is done either by a cell or by something a cell has created. Understanding the organelles and their functions helps us understand how cells are able to create and carry out these complex functions at the organ and organ system levels.

Organelles can be divided into two groups: membranous organelles, those that are surrounded by membrane material, and non-membranous organelles, those that are not surrounded by membrane material. The membranes that surround the membranous organelles and the cell membrane are interchangeable. This interchangeability is an important feature for most of the membranous organelles as membrane segments from one organelle become vesicle membranes that move to and incorporate with the membrane of another organelle during the course of normal cellular activities. The exception are the membranes of mitochondria which are not shared with the other organelles. The collection of organelles that share membrane segments is called the endomembrane system.

The Plasma Membrane

Introduction

Despite differences in structure and function, all living cells in multicellular organisms have a surrounding cell membrane. As the outer layer of your skin separates your body from its environment, the cell membrane (also known as the plasma membrane) separates the inner contents of a cell from its exterior environment. The cell membrane provides a protective barrier around the cell and regulates which materials can pass in or out. Without a cell membrane there would be no way to separate the homeostatic stability of an inside **Intracellular fluid (ICF)** from the chemical fluctuations of the outside **Extracellular fluid (ECF)**, and maintaining the dynamic kinetic stability of life would not be possible.

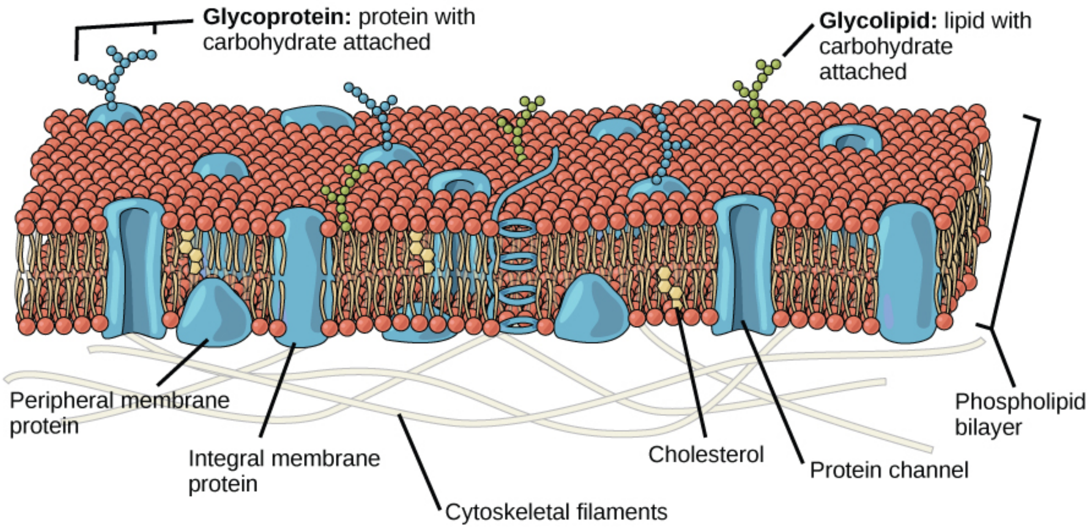
The plasma membrane must be very flexible to allow certain cells, such as red blood cells and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious functions of a plasma membrane. In addition, the surface of the plasma membrane carries markers that allow cells to recognize one another, which is vital for tissue and organ formation during early development, and which later plays a role in the "self" versus "non-self" distinction of the immune response.

Among the most sophisticated functions of the plasma membrane is the ability to transmit signals by means of complex, integral proteins known as **membrane receptors**. These proteins (and occasionally, lipids) act both as receivers of extracellular inputs and as activators of intracellular processes. These membrane receptors provide extracellular attachment sites for effectors like hormones and growth factors, and they activate intracellular response cascades when their effectors are bound. Occasionally, receptors are hijacked by viruses (HIV, human immunodeficiency virus, is one example) that use them to gain entry into cells, and at times, the genes encoding receptors become mutated, causing the process of signal transduction to malfunction with disastrous consequences.

Fluid Mosaic Model

The existence of the plasma membrane was identified in the 1890s, and its chemical components were identified in 1915. The principal components identified at that time were lipids and proteins. The first widely accepted model of the plasma membrane’s structure was proposed in 1935 by Hugh Davson and James Danielli; it was based on the “railroad track” appearance of the plasma membrane in early electron micrographs. They theorized that the structure of the plasma membrane resembles a sandwich, with protein being analogous to the bread, and lipids being analogous to the filling. In the 1950s, advances in microscopy, notably transmission electron microscopy (TEM), allowed researchers to see that the core of the plasma membrane consisted of a double, rather than a single, layer. A new model that better explains both the microscopic observations and the function of that plasma membrane was proposed by S.J. Singer and Garth L. Nicolson in 1972.

The explanation proposed by Singer and Nicolson, and based on the work of many others such as Harden McConnell, is called the **fluid mosaic model**. The model has evolved somewhat over time, but it still best accounts for the structure and functions of the plasma membrane as we now understand them. The fluid mosaic model describes the structure of the plasma membrane as a mosaic of components—including phospholipids, cholesterol, proteins, and carbohydrates—that gives the membrane a fluid character. Plasma membranes range from 5 to 10 nm in thickness. For comparison, human red blood cells, visible via light microscopy, are approximately 8 µm wide, or approximately 1,000 times wider than a plasma membrane. The membrane does look a bit like a sandwich ([link]).


Glycoprotein: protein with carbohydrate attached
Glycolipid: lipid with carbohydrate attached
Phospholipid bilayer
Peripheral membrane protein
Integral membrane protein
Cytoskeletal filaments
Cholesterol
Protein channel

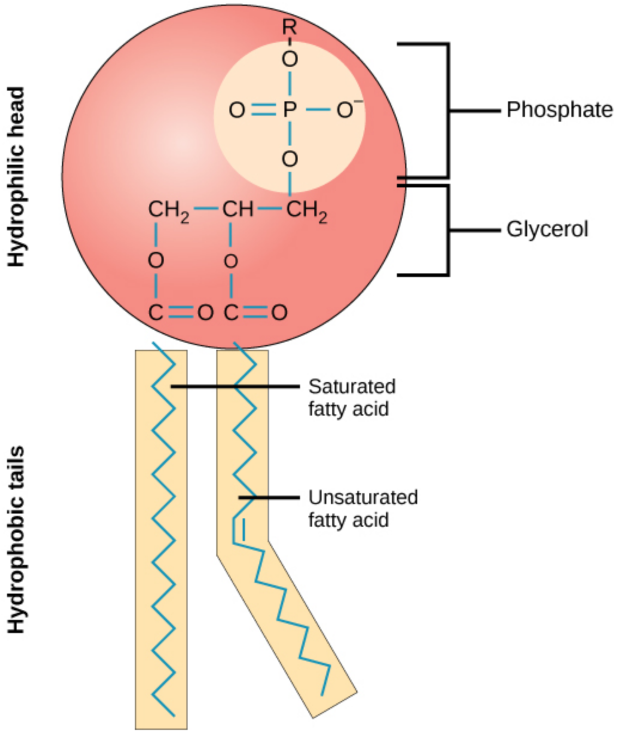
The fluid mosaic model of the plasma membrane describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the outward-facing surface of the membrane.

The principal components of a plasma membrane are lipids (phospholipids and cholesterol), proteins, and carbohydrates attached to some of the lipids and some of the proteins. A **phospholipid** is a molecule consisting of glycerol, two fatty acids, and a phosphate-linked head group. **Cholesterol**, another lipid composed of four fused carbon rings, is found alongside the phospholipids in the core of the membrane. The proportions of proteins, lipids, and carbohydrates in the plasma membrane vary with cell type, but for a typical human cell, protein accounts for about 50 percent of the composition by mass, lipids (of all types) account for about 40 percent of the composition by mass, with the remaining 10 percent of the composition by mass being carbohydrates. However, the concentration of proteins and lipids varies with different cell membranes. For example, myelin, an outgrowth of the membrane of specialized cells that insulates the axons of the peripheral nerves, contains only 18 percent protein and 76 percent lipid. The mitochondrial inner membrane contains 76 percent protein and only 24 percent lipid. The plasma membrane of human red blood cells is 30 percent lipid. Carbohydrates are present only on the exterior surface of the plasma membrane and are attached to proteins, forming glycoproteins, or attached to lipids, forming glycolipids.

Phospholipids

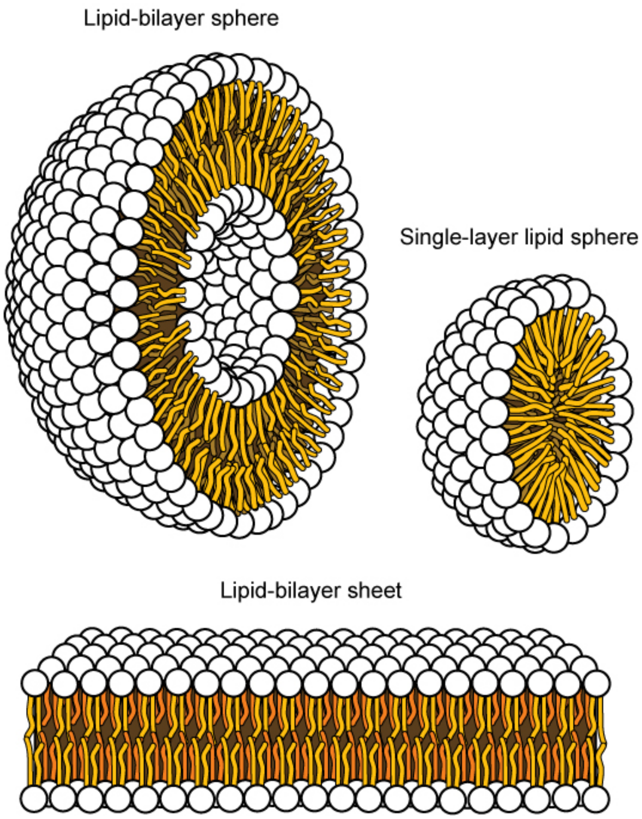
The main fabric of the membrane is composed of amphiphilic, phospholipid molecules. The **hydrophilic** or “water-loving” areas of these molecules (which look like a collection of balls in an artist’s rendition of the model) ([link]) are in contact with the aqueous fluid both inside and outside the cell. **Hydrophobic**, or water-hating molecules, tend to be non-polar. They interact with other non-polar molecules in chemical reactions, but generally do not interact with polar molecules. When placed in water, hydrophobic molecules tend to form a ball or cluster. The hydrophilic regions of the phospholipids tend to form hydrogen bonds with water and other polar molecules on both the exterior and interior of the cell. Thus, the membrane surfaces that face the interior and exterior of the cell are hydrophilic. In contrast, the interior of the cell membrane is hydrophobic and will not interact with water. Therefore, phospholipids form an excellent two-layercell membrane that separates fluid within the cell from the fluid outside of the cell.

A phospholipid molecule ([link]) consists of a three-carbon glycerol backbone with two fatty acid molecules attached to carbons 1 and 2, and a phosphate-containing group attached to the third carbon. This arrangement gives the overall molecule an area described as its head (the phosphate-containing group), which has a polar character or negative charge, and an area called the tail (the fatty acids), which has no charge. The head can form hydrogen bonds, but the tail cannot. A molecule with this arrangement of a positively or negatively charged area and an uncharged, or non-polar, area is referred to as **amphiphilic** or “dual-loving.”



This phospholipid molecule is composed of a hydrophilic head and two hydrophobic tails. The hydrophilic head group consists of a phosphate-containing group attached to a glycerol molecule. The hydrophobic tails, each containing either a saturated or an unsaturated fatty acid, are long hydrocarbon chains.

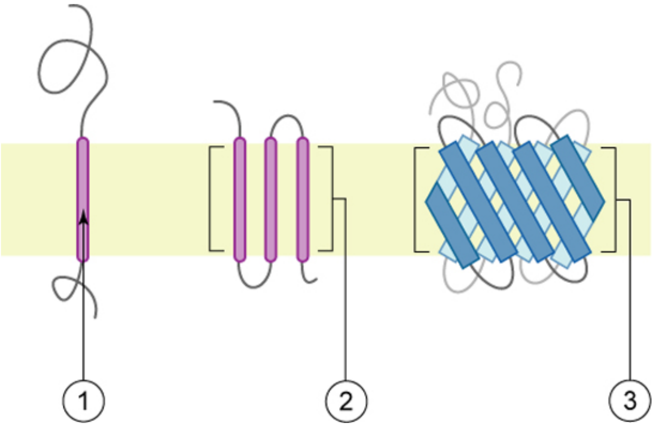
This characteristic is vital to the structure of a plasma membrane because, in water, phospholipids tend to become arranged with their hydrophobic tails facing each other and their hydrophilic heads facing out. In this way, they form a lipid bilayer—a barrier composed of a double layer of phospholipids that separates the water and other materials on one side of the barrier from the water and other materials on the other side. In fact, phospholipids heated in an aqueous solution tend to spontaneously form small spheres or droplets (called micelles or liposomes), with their hydrophilic heads forming the exterior and their hydrophobic tails on the inside ([link]).



In an aqueous solution, phospholipids tend to arrange themselves with their polar heads facing outward and their hydrophobic tails facing inward. (credit: modification of work by Mariana Ruiz Villareal)

Membrane Proteins

Integral proteins (some specialized types are called integrins) are, as their name suggests, integrated completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the hydrophobic region of the the phospholipid bilayer ([link]). Single-pass integral membrane proteins usually have a hydrophobic transmembrane segment that consists of 20–25 amino acids. Some span only part of the membrane—associating with a single layer—while others stretch from one side of the membrane to the other, and are exposed on either side. Some complex proteins are composed of up to 12 segments of a single protein, which are extensively folded and embedded in the membrane ([link]). This type of protein has a hydrophilic region or regions, and one or several mildly hydrophobic regions. This arrangement of regions of the protein tends to orient the protein alongside the phospholipids, with the hydrophobic region of the protein adjacent to the tails of the phospholipids and the hydrophilic region or regions of the protein protruding from the membrane and in contact with the cytosol or extracellular fluid.



Integral membranes proteins may have one or more alpha-helices that span the membrane (examples 1 and 2), or they may have beta-sheets that span the membrane (example 3). (credit: “Foobar”/Wikimedia Commons)

A **channel protein** is an example of an integral protein that selectively allows particular materials, such as certain ions, to pass into or out of the cell.

Another important group of integral proteins are **cell recognition proteins**, which serve to mark a cell's identity so that it can be recognized by other cells. A **receptor** is a type of recognition protein that can selectively bind a specific molecule outside the cell, and this binding induces a chemical reaction within the cell. A **ligand** is the specific molecule that binds to and activates a receptor. Some integral proteins serve dual roles as both a receptor and an ion channel. One example of a receptor-ligand interaction is the receptors on nerve cells that bind neurotransmitters, such as dopamine. When a dopamine molecule binds to a dopamine receptor protein, a channel within the transmembrane protein opens to allow certain ions to flow into the cell.

Some integral membrane proteins are glycoproteins. A **glycoprotein** is a protein that has carbohydrate molecules attached, which extend into the extracellular matrix. The attached carbohydrate tags on glycoproteins aid in cell recognition. The carbohydrates that extend from membrane proteins and even from some membrane lipids collectively form the glycocalyx. The **glycocalyx** is a fuzzy-appearing coating around the cell formed from glycoproteins and other carbohydrates attached to the cell membrane. The glycocalyx can have various roles. For example, it may have molecules that allow the cell to bind to another cell, it may contain receptors for hormones, or it might have enzymes to break down nutrients. The glycocalyces found in a person's body are products of that person's genetic makeup. They give each of the individual's trillions of cells the "identity" of belonging in the person's body. This identity is the primary way that a person's immune defense cells "know" not to attack the person's own body cells, but it also is the reason organs donated by another person might be rejected.

Peripheral proteins are found on the exterior and interior surfaces of membranes, attached either to integral proteins or to phospholipids. Peripheral proteins, along with integral proteins, may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell’s recognition sites. These are sometimes referred to as “cell-specific” proteins. The body recognizes its own proteins and attacks foreign proteins associated with invasive pathogens.

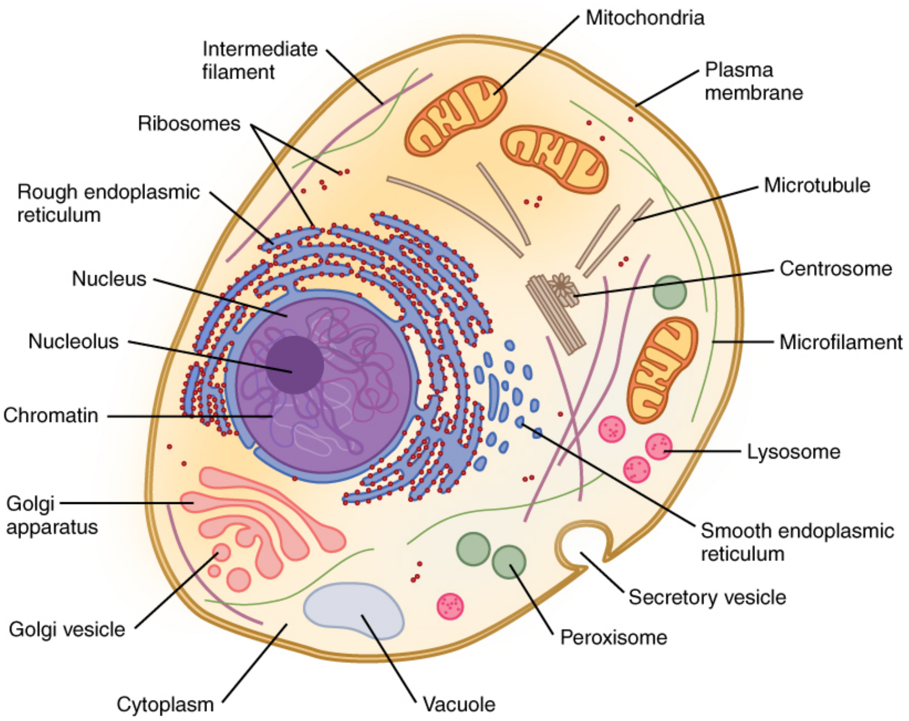
The Cytoplasm and Cellular Organelles

By the end of this section, you will be able to:

- Describe the structure and function of the cellular organelles associated with the endomembrane system, including the endoplasmic reticulum, Golgi apparatus, and lysosomes
- Describe the structure and function of mitochondria and peroxisomes
- Explain the three components of the cytoskeleton, including their composition and functions

Now that you have learned that the cell membrane surrounds all cells, you can dive inside of a prototypical human cell to learn about its internal components and their functions. All living cells in multicellular organisms contain an internal cytoplasmic compartment, and a nucleus within the cytoplasm. **Cytosol**, the jelly-like substance within the cell, provides the fluid medium necessary for biochemical reactions. Eukaryotic cells, including all animal cells, also contain various cellular organelles. An **organelle** (“little organ”) is one of several different types of membrane-enclosed bodies in the cell, each performing a unique function. Just as the various bodily organs work together in harmony to perform all of a human’s functions, the many different cellular organelles work together to keep the cell healthy and performing all of its important functions. The organelles and cytosol, taken together, compose the cell’s **cytoplasm**. The **nucleus** is a cell’s central organelle, which contains the cell’s DNA ([link]).

Prototypical Human Cell



While this image is not indicative of any one particular human cell, it is a prototypical example of a cell containing the primary organelles and internal structures.

The Nucleus and Nuclear Structures

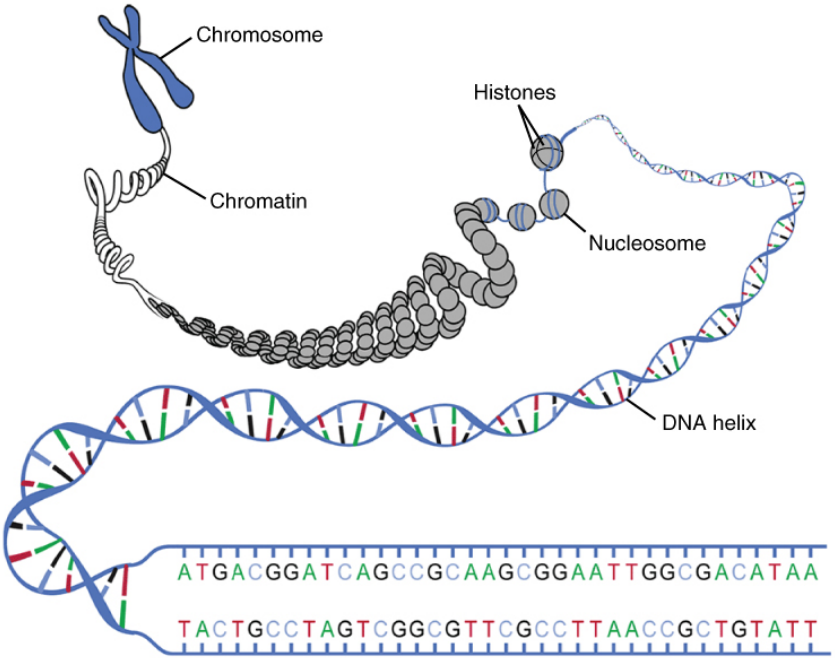
The nucleus is the largest and most prominent of a cell’s organelles ([link]). The nucleus is generally considered the control center of the cell because it stores all of the genetic instructions for manufacturing proteins.

Like most other cellular organelles, the nucleus is surrounded by a membrane called the **nuclear envelope**. This membranous covering consists of two adjacent lipid bilayers with a thin fluid space in between them. Spanning these two bilayers are nuclear pores. A **nuclear pore** is a tiny passageway for the passage of proteins, RNA, and solutes between the nucleus and the cytoplasm. Proteins called pore complexes lining the nuclear pores regulate the passage of materials into and out of the nucleus.

Inside the nuclear envelope is a gel-like nucleoplasm with solutes that include the building blocks of nucleic acids. There also can be a dark-staining mass often visible under a simple light microscope, called a **nucleolus** (plural = nucleoli). The nucleolus is a region of the nucleus that is responsible for manufacturing the RNA necessary for construction of ribosomes. Once synthesized, newly made ribosomal subunits exit the cell's nucleus through the nuclear pores.

The genetic instructions that are used to build and maintain an organism are arranged in an orderly manner in strands of DNA. Within the nucleus are threads of **chromatin** composed of DNA and associated proteins ([link]). Along the chromatin threads, the DNA is wrapped around a set of **histone** proteins. A **nucleosome** is a single, wrapped DNA-histone complex. Multiple nucleosomes along the entire molecule of DNA appear like a beaded necklace, in which the string is the DNA and the beads are the associated histones. When a cell is in the process of division, the chromatin condenses into chromosomes, so that the DNA can be safely transported to the "daughter cells." The **chromosome** is composed of DNA and proteins; it is the condensed form of chromatin.

DNA Macrostructure


Chromosome
Histones
Chromatin
Nucleosome
DNA helix
ATGACGGATCAGCCGCAAGCGGAATTGGCGACATAA
TACTGCCTAGTCGGCGTTCGCCTTAACCGCTGTATT

Strands of DNA are wrapped around supporting histones. These proteins are

increasingly bundled and condensed into
chromatin, which is packed tightly into
chromosomes when the cell is ready to divide.

It is estimated that humans have almost 22,000 genes distributed on 46 chromosomes. These **genes** (nucleotide sequences that code for proteins) account for only a small percentage of the nucleotides making up the chromosomes. Many of the remaining nucleotides are transcribed into RNA molecules called **long non-coding RNA** (lncRNA) that are important for regulating gene expression and for creating and maintaining the 3-dimesional structure of the nucleus.

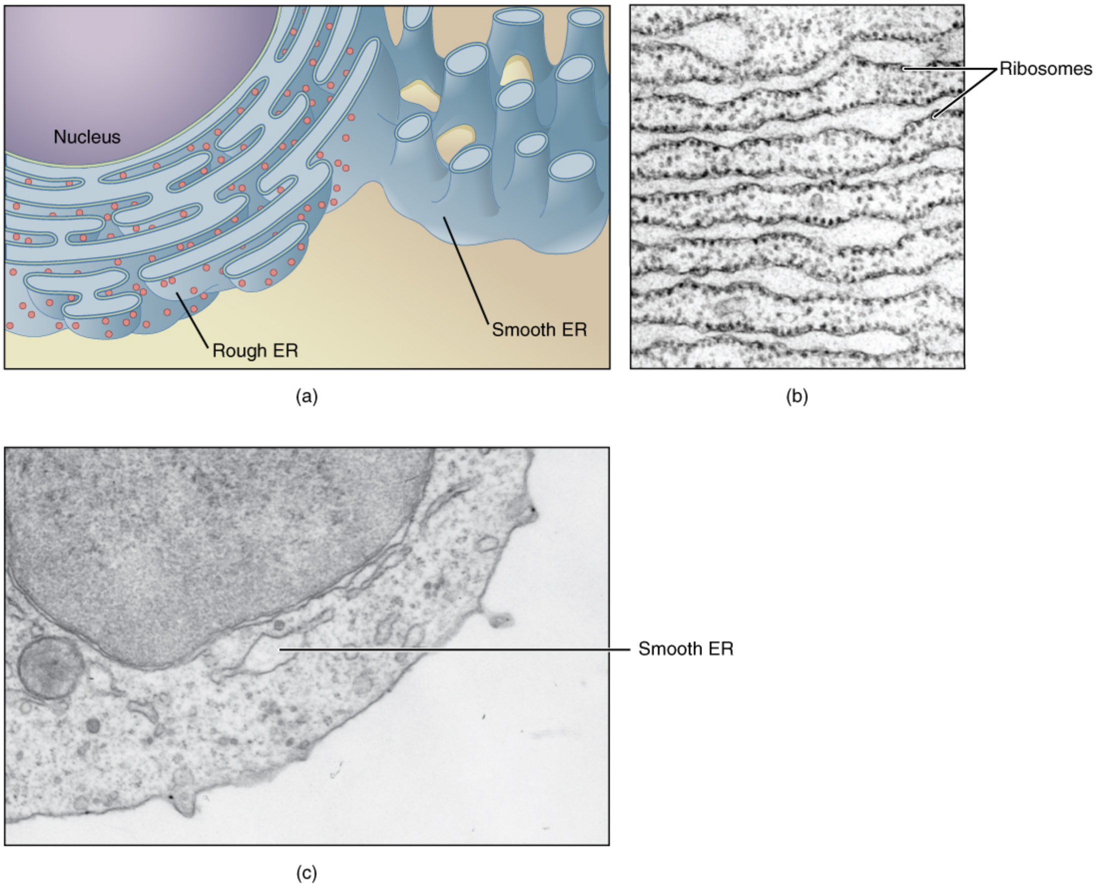
Organelles of the Endomembrane System

A set of three major kinds of organelles together form a system within the cell called the endomembrane system. These organelles work together to perform various cellular jobs, including the task of producing, packaging, and exporting certain cellular products. The organelles of the endomembrane system include the endoplasmic reticulum, Golgi apparatus, and vesicles. These organelles all "share" pieces of their membrane. This occurs when part of an organelle's membrane pinches off to form a vesicle membrane. When that vesicle fuses with another organelle (or the cell membrane) the pinched-off piece of membrane becomes part of the next organelle's membrane. Thus the organelles of the endomembrane system are continually swapping pieces of their membrane.

Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** is a system of channels that is continuous with the nuclear membrane (or “envelope”) covering the nucleus and composed of the same lipid bilayer material. The ER can be thought of as a series of winding thoroughfares similar to the waterway canals in Venice. The ER provides passages throughout much of the cell that function in transporting, synthesizing, and storing materials. The winding structure of the ER results in a large membranous surface area that supports its many functions ([link]). Endoplasmic reticulum can exist in two forms: rough ER and smooth ER. These two types of ER perform some very different functions and can be found in very different amounts depending on the type of cell.

Endoplasmic Reticulum (ER)



(a) The ER is a winding network of thin membranous sacs found in close association with the cell nucleus. The smooth and rough endoplasmic reticula are very different in appearance and function (source: mouse tissue). (b) Rough ER is studded with numerous ribosomes, which are sites of protein synthesis (source: mouse tissue). EM × 110,000. (c) Smooth ER synthesizes phospholipids, steroid hormones, regulates the concentration of cellular Ca^{++}, metabolizes some carbohydrates, and breaks down certain toxins (source: mouse tissue). EM × 110,510. (Micrographs provided by the Regents of University of Michigan Medical School © 2012)

Rough ER (RER) is so-called because its membrane is dotted with embedded non-membranous organelles called ribosomes, giving the RER a "rough" appearance. **Ribosomes** are programmable enzymatic workbenches that serve as the site of protein synthesis. They are composed of two ribosomal RNA subunits that wrap around a messenger RNA (mRNA) molecule to start the process of translating the nucleotide sequence into the corresponding amino acid sequence. The mRNA sequence provides the ribosome with the "program" to make a specific protein.

The primary function of the rough ER is the synthesis and modification of proteins destined for the cell membrane or for export from the cell. For this protein synthesis, many ribosomes attach to the ER (giving it the studded appearance of rough ER). Typically, a protein is synthesized within the ribosome and released through a channel to a cistern (chamber) inside the rough ER, where sugars can be added to it (by a process called glycosylation). The contents are enclosed in a vesicle that forms by pinching off part of the ER membrane. The contents are transported within the vesicle to the next stage in the packaging and shipping process: the Golgi apparatus.

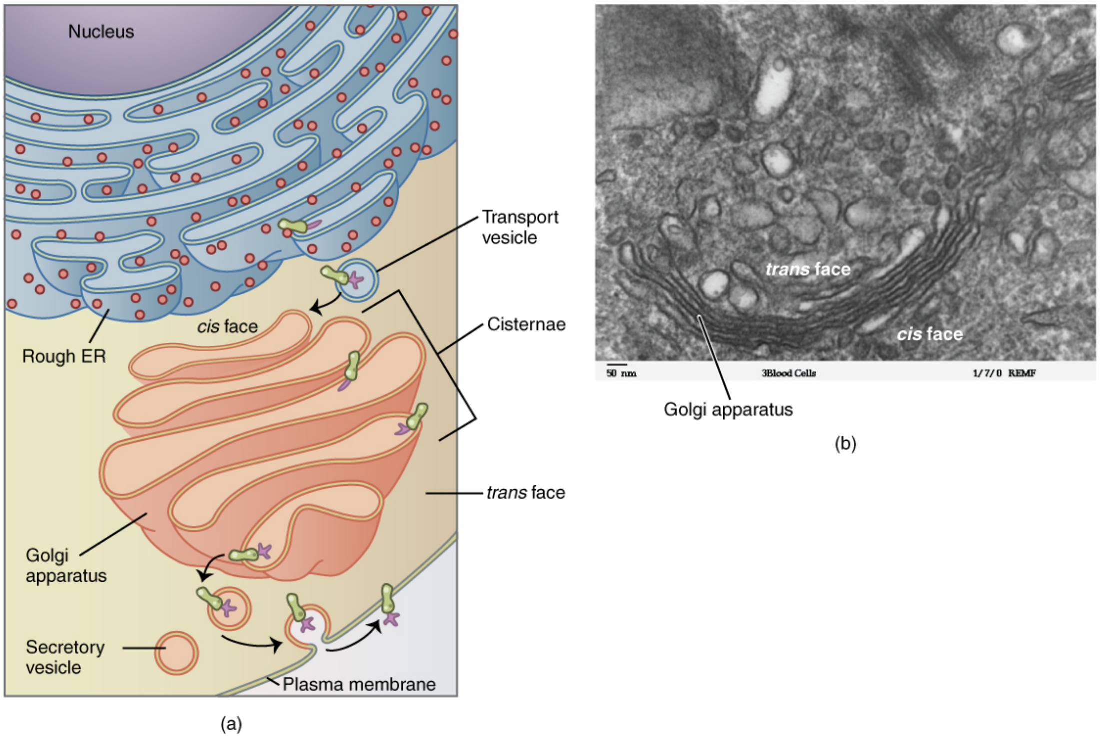
Smooth ER (SER) lacks attached ribosomes but contains enzymes that perform specific functions. The enzymes vary with cell type and give many cells and their organs their specific functionality. For example, smooth ER is important in the synthesis of lipids. It is the site where phospholipids, the main component of biological membranes, are manufactured. Enzymes found in the SER also are important in the detoxification of certain chemicals and the metabolism of some carbohydrates. Because many of these functions occur in the liver, the liver contains large amounts of smooth ER.

The smooth ER also houses enzymes for steroid hormone manufacture. For this reason, cells that produce large quantities of such hormones, such as those of the female ovaries and male testes, contain large amounts of smooth ER. In addition to lipid synthesis, the smooth ER also sequesters (i.e., stores) and regulates the concentration of cellular Ca^{++}, which is important for many metabolic functions. For example, in skeletal muscle cells Ca^{++} is the trigger for contractions. The smooth ER in skeletal muscle acts as an internal Ca^{++} container supporting the rapid release and re-uptake of Ca^{++} through specialized channels in the SER membrane.

The Golgi Apparatus

The **Golgi apparatus** is responsible for sorting, modifying, and shipping off the products that come from the rough ER, much like a post-office. The Golgi apparatus looks like stacked flattened discs, almost like stacks of oddly shaped pancakes. Like the ER, these discs are membranous. The Golgi apparatus has two distinct sides, each with a different role. One side of the apparatus receives products in vesicles. These products are sorted through the apparatus, and then they are released from the opposite side after being repackaged into new vesicles. If the product is to be exported from the cell, the vesicle migrates to the cell surface and fuses to the cell membrane, and the cargo is secreted ([link]).

Golgi Apparatus



(a) The Golgi apparatus manipulates products from the rough ER, and also produces new organelles called lysosomes. Proteins and other products of the ER are sent to the Golgi apparatus, which organizes, modifies, packages, and tags them. Some of these products are transported to other areas of the cell and some are exported from the cell through exocytosis. Enzymatic proteins are packaged as new lysosomes (or packaged and sent for fusion with existing lysosomes). (b) An electron micrograph of the Golgi apparatus.

Lysosomes

Some of the protein products packaged by the Golgi include digestive enzymes that are meant to remain inside the cell for use in breaking down certain materials. The enzyme-containing vesicles released by the Golgi may form new lysosomes, or fuse with existing, lysosomes. A **lysosome** is an organelle that contains enzymes that break down and digest unneeded cellular components, such as a damaged organelle. (A lysosome is similar to a wrecking crew that takes down old and unsound buildings in a neighborhood.) **Autophagy** (“self-eating”) is the process of a cell digesting its own structures. Lysosomes are also important for breaking down foreign material. For example, when certain immune defense cells (white blood cells) phagocytize bacteria, the bacterial cell is transported into a lysosome and digested by the enzymes inside. As one might imagine, such phagocytic defense cells contain large numbers of lysosomes.

Under certain circumstances, lysosomes perform a more grand and dire function. In the case of damaged or unhealthy cells, lysosomes can be triggered to open up and release their digestive enzymes into the cytoplasm of the cell, killing the cell. This “self-destruct” mechanism is called **autolysis**, and makes the process of cell death controlled (a mechanism called “apoptosis”).

Note:

Watch this video to learn about the endomembrane system, which includes the rough and smooth ER and the Golgi body as well as lysosomes and vesicles. What is the primary role of the endomembrane system?

Channel and Transporter Protein-Storage Vesicles

In addition to using vesicles to separate reactive solutions and enzymes from the cytoplasm during transport, cells also use vesicles to store channel and transporter proteins when they are not being used at the cell surface. The proteins are stored as part of the vesicle membrane in the cytoplasm. When the proteins are needed the vesicles translocate to the surface where they fuse with the cell membrane. This integrates the vesicle membrane containing the channels or transporters into the cell membrane making them functional. When they are no longer needed those portions of the cell membrane are pinched off to reform as vesicles in the cytoplasm. This is an efficient way to rapidly manage cellular interactions with the external environment.

The translocation and fusion processes, and their reverse, are often controlled by hormones. For example, this is how insulin stimulates muscle and fat cells to remove glucose from circulation. The insulin binds to surface receptors which begin a cascade of events in the cell. One of the results is the translocation of vesicles with GLUT4 carrier proteins embedded in their membranes. When the vesicles fuse with the cell membrane the GLUT4 carrier proteins can begin transporting glucose into the cell. A similar process is used to manage water retention in the kidneys. Antidiuretic Hormone (ADH), also known as Vasopressin, stimulates translocation of vesicles containing Aquaporin (AQP2) channels to the tubule cell surface allowing water to be reabsorbed.

Membranous Organelles for Detoxification and Energy Production

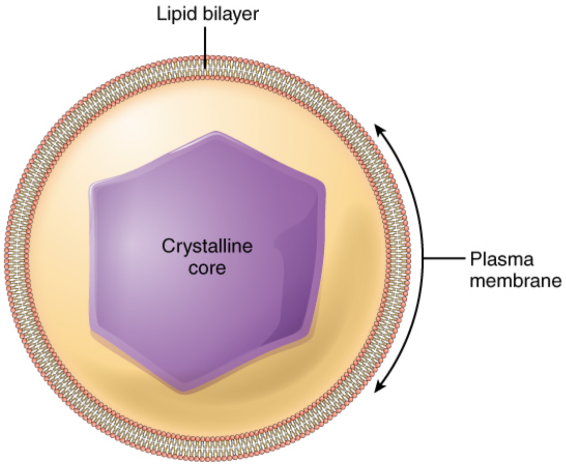
In addition to the jobs outlined above the cell has many other important functions. One important function of the cell is detoxification. Humans take in all sorts of toxins from the environment and also produce harmful chemicals as byproducts of cellular processes. Cells called hepatocytes in the liver use peroxisomes to detoxify many of these toxins.

You must consume nutrients to provide the energy required to maintain the dynamic kinetic stability of your cells and organs. The nutrients you consume must be converted to chemical energy in the form of ATP in order to power these biochemical reactions. While small amounts of ATP can be produced by enzymes in the cytoplasm (about 5% of the total that can be extracted from a molecule of glucose) the remaining 95% is produced in specialized organelles called mitochondria.

Peroxisomes

Like lysosomes, a **peroxisome** is a membrane-bound cellular organelle that contains mostly enzymes ([link]). Peroxisomes, a product of the endoplasmic reticulum, perform a couple of different functions, including lipid metabolism and chemical detoxification. In contrast to the digestive enzymes found in lysosomes, the enzymes within peroxisomes serve to transfer hydrogen atoms from various molecules to oxygen, producing hydrogen peroxide (H_2O_2). In this way, peroxisomes neutralize poisons such as alcohol. In order to appreciate the importance of peroxisomes, it is necessary to understand the concept of reactive oxygen species.

Peroxisome



Peroxisomes are membrane-bound organelles that contain an abundance of enzymes for detoxifying harmful substances and lipid metabolism.

Reactive oxygen species (ROS) such as peroxides and free radicals are the highly reactive products of many normal cellular processes, including the mitochondrial reactions that produce ATP and oxygen metabolism. Examples of ROS include the hydroxyl radical OH, H_2O_2, and superoxide (O_2^-). Some ROS are important for certain cellular functions, such as cell signaling processes and immune responses against foreign substances. Free radicals are reactive because they contain free unpaired electrons; they can easily oxidize other molecules throughout the cell, causing cellular damage and even cell death. Free radicals are thought to play a role in many destructive processes in the body, from cancer to coronary artery disease.

Peroxisomes, on the other hand, oversee reactions that neutralize free radicals. Peroxisomes produce large amounts of the toxic H_2O_2 in the process, but peroxisomes contain enzymes that convert H_2O_2 into water and oxygen. These byproducts are safely released into the cytoplasm. Like miniature sewage treatment plants, peroxisomes neutralize harmful toxins so that they do not wreak havoc in the cells. The liver is the organ primarily responsible for detoxifying the blood before it travels throughout the body, and liver cells contain an exceptionally high number of peroxisomes.

Defense mechanisms such as detoxification within the peroxisome and certain cellular antioxidants serve to neutralize many of these molecules. Some vitamins and other substances, found primarily in fruits and vegetables, have antioxidant properties. Antioxidants work by being oxidized themselves, halting the destructive reaction cascades initiated by the free radicals. Sometimes though, ROS accumulate beyond the capacity of such defenses.

Oxidative stress is the term used to describe damage to cellular components caused by ROS. Due to their characteristic unpaired electrons, ROS can set off chain reactions where they remove electrons from other molecules, which then become oxidized and reactive, and do the same to other molecules, causing a chain reaction. ROS can cause permanent damage to cellular lipids, proteins, carbohydrates, and nucleic acids. Damaged DNA can lead to genetic mutations and even cancer. A **mutation** is a change in the nucleotide sequence in a gene within a cell's DNA, potentially altering the protein coded by that gene. Other diseases believed to be triggered or exacerbated by ROS include Alzheimer's disease, cardiovascular diseases, diabetes, Parkinson's disease, arthritis, Huntington's disease, and schizophrenia, among many others. It is noteworthy that these diseases are largely age-related. Many scientists believe that oxidative stress is a major contributor to the aging process.

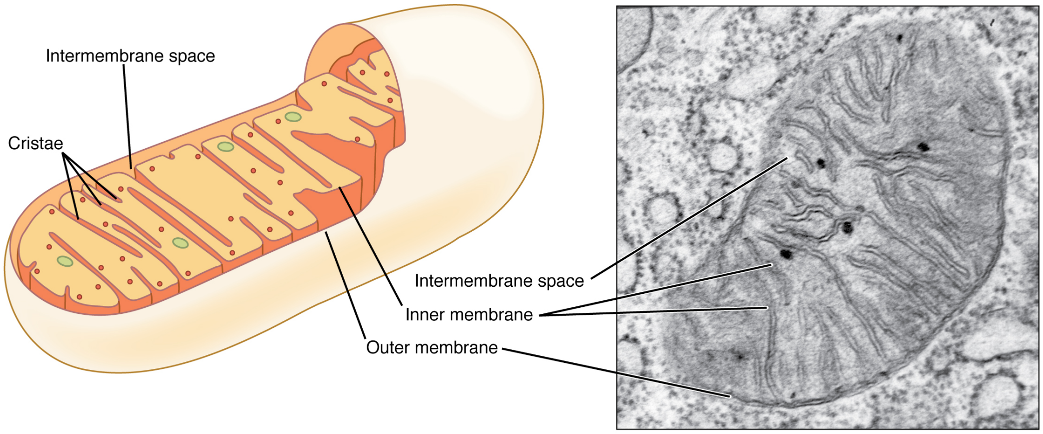
Note:
Free Radicals and Aging
Cell: The Free Radical Theory

The free radical theory on aging was originally proposed in the 1950s, and still remains under debate. Generally speaking, the free radical theory of aging suggests that accumulated cellular damage from oxidative stress contributes to the physiological and anatomical effects of aging. There are two significantly different versions of this theory: one states that the aging process itself is a result of oxidative damage, and the other states that oxidative damage causes age-related disease and disorders. The latter version of the theory is more widely accepted than the former. However, many lines of evidence suggest that oxidative damage does contribute to the aging process. Research has shown that reducing oxidative damage can result in a longer lifespan in certain organisms such as yeast, worms, and fruit flies. Conversely, increasing oxidative damage can shorten the lifespan of mice and worms. Interestingly, a manipulation called calorie-restriction (moderately restricting the caloric intake) has been shown to increase life span in some laboratory animals. It is believed that this increase is at least in part due to a reduction of oxidative stress. However, a long-term study of primates with calorie-restriction showed no increase in their lifespan. A great deal of additional research will be required to better understand the link between reactive oxygen species and aging.

Mitochondria

A **mitochondrion** (plural = mitochondria) is a membranous, bean-shaped organelle that is the “energy transformer” of the cell. Mitochondria have two membranes an outer lipid bilayer membrane as well as an additional inner lipid bilayer membrane ([link]). The inner membrane is highly folded into winding structures with a great deal of surface area, called cristae. It is along this inner membrane that a series of proteins, enzymes, and other molecules perform the biochemical reactions of cellular respiration. These reactions convert energy stored in nutrient molecules (such as glucose) into adenosine triphosphate (ATP), which provides usable cellular energy to the cell. Cells use ATP constantly, and so the mitochondria are constantly at work. Oxygen molecules are required as the final electron acceptor during mitochondrial respiration, which is why you must constantly breathe it in. One of the organ systems in the body that uses huge amounts of ATP is the muscular system because ATP is required to sustain muscle contraction. As a result, muscle cells are packed full of mitochondria. Nerve cells also need large quantities of ATP to run their sodium-potassium pumps. Therefore, an individual neuron will be loaded with over a thousand mitochondria. On the other hand, a bone cell, which is not nearly as metabolically-active, might only have a couple hundred mitochondria.

Mitochondrion



(a) (b)

The mitochondria are the energy-conversion factories of the cell. (a) A mitochondrion is composed of two separate lipid bilayer membranes. Along the inner membrane are various molecules that work together to produce ATP, the cell’s major energy currency. (b) An electron micrograph of mitochondria. EM × 236,000. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

Mitochondria differ from other membranous organelles in that they do not participate in the endomembrane system and they are not produced by either the endoplasmic reticulum like peroxisomes or the Golgi apparatus like lysosomes. Instead mitochondria have their own circular ring of DNA, reproduce by fission in the same way as prokaryotes, like bacteria, and their ribosomes more closely resemble bacterial ribosomes than eukaryotic ribosomes. If the mitochondria are removed from a cell, the cell cannot make new mitochondria because the nuclear DNA does not contain instructions for mitochondria. A consequence of this is that all of your mitochondria came from your mother, because the ovum (egg) is the cell that supplies the cytoplasm and organelles during fertilization.

The resemblance of mitochondria to bacteria has led to the hypothesis called endosymbiosis. It postulates that mitochondria were once free living cells that over one billion years ago became engulfed in ancestral eukaryotic cells. The waste products (pyruvate) of the host cell's anaerobic respiration provided a ready food source for the mitochondria, and the excess ATP the mitochondria produced via oxidative pyruvate metabolism provided an abundant new energy source for the host cell. This type of endosymbiosis is not unique to mitochondria. The chloroplasts in plants and blue-green algae have the same characteristics as mitochondria. And, an analogous kind of endosymbiosis also occurs between fungi and blue-green algae to form lichens.

Non-membranous Organelles

Ribosomes

A **ribosome** is a complex macromolecule composed of structural and catalytic ribosomal RNAs (rRNAs), and many distinct polypeptides. In eukaryotes, like human cells, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

In eukaryotic cells, ribosomes exist both as free organelles in the cytoplasm and attached to receptors on the rough endoplasmic reticulum membrane. Mitochondria have only free ribosomes, which resemble prokaryotic ribosomes (and have similar drug sensitivities).

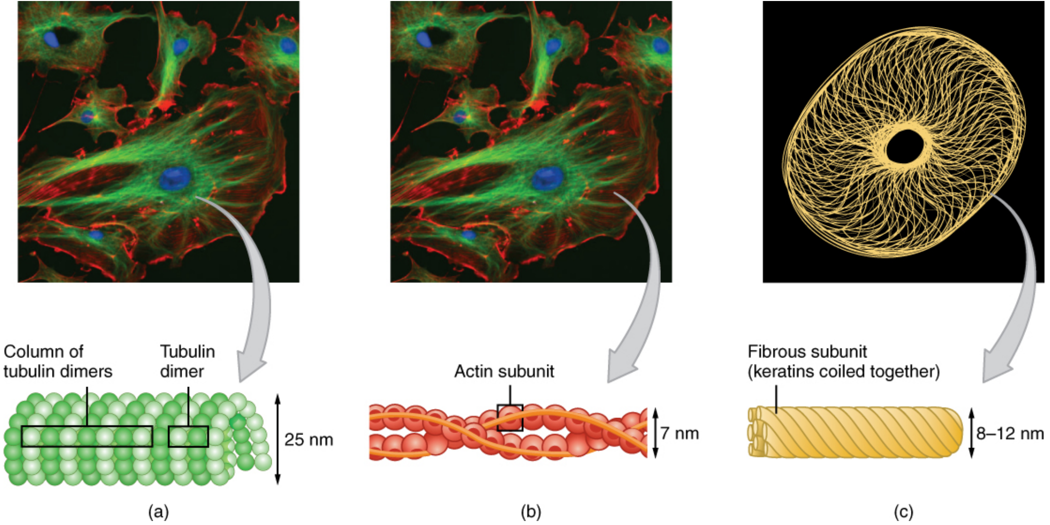
Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds tRNAs. Each mRNA molecule is translated by many ribosomes simultaneously, all synthesizing protein in the same direction: reading the mRNA from the 5' carbon end to 3' carbon end and synthesizing the polypeptide from the N terminus to the C terminus. The complete mRNA/poly-ribosome structure is called a polysome.

The Cytoskeleton

Much like the bony skeleton structurally supports the human body, the cytoskeleton helps the cells to maintain their structural integrity. The **cytoskeleton** is a group of fibrous proteins that provide structural support for cells, but this is only one of the functions of the cytoskeleton. Cytoskeletal components are also critical for cell motility, cell reproduction, and transportation of substances within the cell.

The cytoskeleton forms a complex thread-like network throughout the cell consisting of three different kinds of protein-based filaments: microfilaments, intermediate filaments, and microtubules ([link]). The thickest of the three is the **microtubule**, a structural filament composed of subunits of a protein called tubulin. Microtubules maintain cell shape and structure, help resist compression of the cell, and play a role in positioning the organelles within the cell. Microtubules also make up two types of cellular appendages important for motion: cilia and flagella. **Cilia** are found on many cells of the body, including the epithelial cells that line the airways of the respiratory system. Cilia move rhythmically; they beat constantly, moving waste materials such as dust, mucus, and bacteria upward through the airways, away from the lungs and toward the mouth. Beating cilia on cells in the female fallopian tubes move egg cells from the ovary towards the uterus. A **flagellum** (plural = flagella) is an appendage larger than a cilium and specialized for cell locomotion. The only flagellated cell in humans is the sperm cell that must propel itself towards female egg cells.

The Three Components of the Cytoskeleton



The cytoskeleton consists of (a) microtubules, (b) microfilaments, and (c) intermediate filaments. The cytoskeleton plays an important role in maintaining cell shape and structure, promoting cellular movement, and aiding cell division.

A very important function of microtubules is to set the paths (somewhat like railroad tracks) between intracellular locations. Special molecules that change shape when they interact with ATP attach to internal structures (like vesicles and chromosomes) and "walk" them down the microtubule to their proper destinations. Two short, identical microtubule structures called centrioles are found near the nucleus of cells. In some processes a **centriole** can serve as the cellular origin point for microtubules extending outward as cilia, flagella, or the mitotic spindles that assist with the separation of DNA during cell division. Microtubules grow out from the centrioles by adding more tubulin subunits, like adding additional links to a chain.

In contrast with microtubules, the **microfilament** is a thinner type of cytoskeletal filament (see [link]**b**). Actin, a protein that forms chains, is the primary component of these microfilaments. Actin fibers, twisted chains of actin filaments, constitute a large component of muscle tissue and, along with the intermediate filament protein, myosin, is responsible for muscle contraction. Like microtubules, actin filaments are long chains of single subunits (called actin subunits). In muscle cells, these long actin strands, called thin filaments, are “pulled” by thick filaments of the myosin protein to contract the cell. Note: the term thick filament here refers to its relative thickness compared only to the actin myofibrils. As a cytoskeletal component myosin is considered an intermediate filament.

Actin also has an important role during cell division. When a cell is about to split in half during cell division, actin filaments work with myosin to create a cleavage furrow that eventually splits the cell down the middle, forming two new cells from the original cell.

The final cytoskeletal filament is the intermediate filament. As its name would suggest, an **intermediate filament** is a filament intermediate in thickness between the microtubules and microfilaments (see [link]**c**). Intermediate filaments are made up of long fibrous subunits. The two most common intermediate filaments are the proteins keratin and myosin. Keratin fibers are wound together like the threads that compose a rope. Intermediate filaments, in concert with the microtubules, are important for maintaining cell shape and structure. Unlike the microtubules, which resist compression, intermediate filaments resist tension—the forces that pull apart cells. There are many cases in which cells are prone to tension, such as when epithelial cells of the skin are compressed, tugging them in different directions. Intermediate filaments help anchor organelles together within a cell and also link cells to other cells by forming special cell-to-cell junctions.

Chapter Review

The internal environmental of a living cell is made up of a fluid, jelly-like substance called cytosol, which consists mainly of water, but also contains various dissolved nutrients and other molecules. The cell is filled with an array of cellular organelles, each one performing a unique function and helping to maintain the health and activity of the cell. The cytosol and organelles together compose the cell’s cytoplasm. Most organelles are surrounded by a lipid membrane similar to the cell membrane of the cell. The endoplasmic reticulum (ER), Golgi apparatus, and lysosomes share a functional connectivity and are collectively referred to as the endomembrane system. There are two types of ER: smooth and rough. While the smooth ER performs many functions, including lipid synthesis and ion storage, the rough ER is mainly responsible for protein synthesis using its associated ribosomes. The rough ER sends newly made proteins to the Golgi apparatus where they are modified and packaged for delivery to various locations within or outside of the cell. Some of these protein products are enzymes destined to break down unwanted material and are packaged as lysosomes for use inside the cell.

Cells also contain peroxisomes and mitochondria, which are the organelles responsible for detoxifying certain chemicals and producing the cell’s energy supply, respectively. Peroxisomes contain enzymes that transform harmful substances such as free radicals into oxygen and water. Biochemical reactions within mitochondria transform energy-carrying molecules into the usable form of cellular energy known as ATP. Cells also contain a miniaturized “skeleton” of protein filaments that extend throughout its interior. Three different kinds of filaments compose this cytoskeleton (in order of increasing thickness): microfilaments, intermediate filaments, and microtubules. Each cytoskeletal component performs unique functions as well as provides a supportive framework for the cell.

Interactive Link Questions

Exercise:

Problem:

Watch this video to learn about the endomembrane system, which includes the rough and smooth ER and the Golgi body as well as lysosomes and vesicles. What is the primary role of the endomembrane system?

Solution:

Processing, packaging, and moving materials manufactured by the cell.

Review Questions

Exercise:

Problem:

Choose the term that best completes the following analogy: Cytoplasm is to cytosol as a swimming pool containing chlorine and flotation toys is to ________.

a. the walls of the pool
b. the chlorine

c. the flotation toys
d. the water

Solution:

D

Exercise:

Problem:

The rough ER has its name due to what associated structures?

a. Golgi apparatus
b. ribosomes
c. lysosomes
d. proteins

Solution:

B

Exercise:

Problem: Which of the following is a function of the rough ER?

a. production of proteins
b. detoxification of certain substances
c. synthesis of steroid hormones
d. regulation of intracellular calcium concentration

Solution:

A

Exercise:

Problem:

Which of the following is a feature common to all three components of the cytoskeleton?

a. They all serve to scaffold the organelles within the cell.
b. They are all characterized by roughly the same diameter.
c. They are all polymers of protein subunits.
d. They all help the cell resist compression and tension.

Solution:

C

Exercise:

Problem:

Which of the following organelles produces large quantities of ATP when both glucose and oxygen are available to the cell?

a. mitochondria
b. peroxisomes
c. lysosomes
d. ER

Solution:

A

Critical Thinking Questions

Exercise:

Problem:

Explain why the structure of the ER, mitochondria, and Golgi apparatus assist their respective functions.

Solution:

The structure of the Golgi apparatus is suited to its function because it is a series of flattened membranous discs; substances are modified and packaged in sequential steps as they travel from one disc to the next. The structure of Golgi apparatus also involves a receiving face and a sending face, which organize cellular products as they enter and leave the Golgi apparatus. The ER and the mitochondria both have structural specializations that increase their surface area. In the mitochondria, the inner membrane is extensively folded, which increases surface area for ATP production. Likewise, the ER is elaborately wound throughout the cell, increasing its surface area for functions like lipid synthesis, Ca^{++} storage, and protein synthesis.

Exercise:

Problem:

Compare and contrast lysosomes with peroxisomes: name at least two similarities and one difference.

Solution:

Peroxisomes and lysosomes are both cellular organelles bound by lipid bilayer membranes, and they both contain many enzymes. However, peroxisomes contain enzymes that detoxify substances by transferring hydrogen atoms and producing H_2O_2, whereas the enzymes in lysosomes function to break down and digest various unwanted materials.

References

Kolata, G. Severe diet doesn’t prolong life, at least in monkeys. *New York Times* [Internet]. 2012 Aug. 29 [cited 2013 Jan 21]; Available from:

http://www.nytimes.com/2012/08/30/science/low-calorie-diet-doesnt-prolong-life-study-of-monkeys-finds.html?_r=2&ref=caloricrestriction&

Glossary

autolysis
breakdown of cells by their own enzymatic action

autophagy
lysosomal breakdown of a cell’s own components

centriole
small, self-replicating organelle that provides the origin for microtubule growth and moves DNA during cell division

cilia
small appendage on certain cells formed by microtubules and modified for movement of materials across the cellular surface

cytoplasm
internal material between the cell membrane and nucleus of a cell, mainly consisting of a water-based fluid called cytosol, within which are all the other organelles and cellular solute and suspended materials

cytoskeleton
“skeleton” of a cell; formed by rod-like proteins that support the cell’s shape and provide, among other functions, locomotive abilities

cytosol
clear, semi-fluid medium of the cytoplasm, made up mostly of water

endoplasmic reticulum (ER)
cellular organelle that consists of interconnected membrane-bound tubules, which may or may not be associated with ribosomes (rough type or smooth type, respectively)

flagellum
appendage on certain cells formed by microtubules and modified for movement

Golgi apparatus
cellular organelle formed by a series of flattened, membrane-bound sacs that functions in protein modification, tagging, packaging, and transport

intermediate filament
type of cytoskeletal filament made of keratin, characterized by an intermediate thickness, and playing a role in resisting cellular tension

lysosome
membrane-bound cellular organelle originating from the Golgi apparatus and containing digestive enzymes

microfilament
the thinnest of the cytoskeletal filaments; composed of actin subunits that function in muscle contraction and cellular structural support

microtubule
the thickest of the cytoskeletal filaments, composed of tubulin subunits that function in cellular movement and structural support

mitochondrion
one of the cellular organelles bound by a double lipid bilayer that function primarily in the production of cellular energy (ATP)

mutation
change in the nucleotide sequence in a gene within a cell’s DNA

nucleus
cell’s central organelle; contains the cell’s DNA

organelle
any of several different types of membrane-enclosed specialized structures in the cell that perform specific functions for the cell

peroxisome

membrane-bound organelle that contains enzymes primarily responsible for detoxifying harmful substances

reactive oxygen species (ROS)

a group of extremely reactive peroxides and oxygen-containing radicals that may contribute to cellular damage

ribosome

cellular organelle that functions in protein synthesis

Connections Between Cells and Cellular Activities

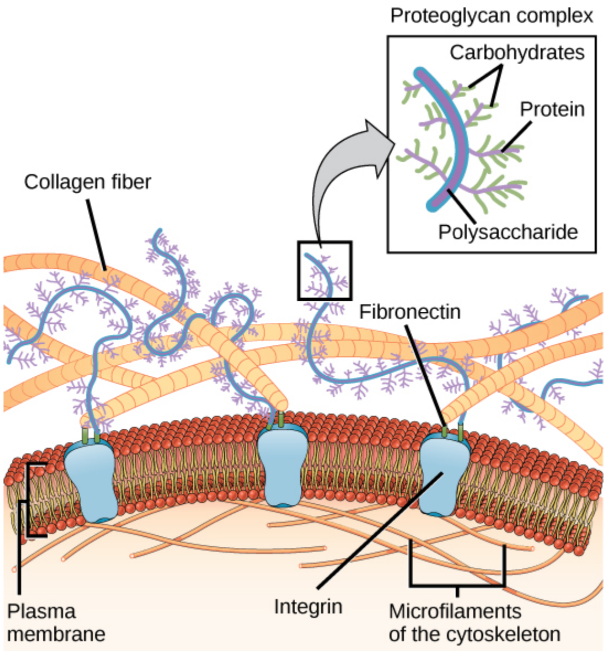
By the end of this section, you will be able to:

- Describe the extracellular matrix
- List examples of the ways that animal cells communicate with adjacent cells
- Summarize the roles of tight junctions, desmosomes, and gap junctions

You already know that a group of similar cells working together is called a tissue. As you might expect, if cells are to work together, they must communicate with each other, just as you need to communicate with others if you work on a group project. Let’s take a look at how cells communicate with each other.

Extracellular Matrix of Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are proteins, and the most abundant protein is collagen. Collagen fibers are interwoven with carbohydrate-containing protein molecules called proteoglycans. Collectively, these materials are called the **extracellular matrix** ([link]). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other. How can this happen?



The extracellular matrix consists of a network of proteins and carbohydrates.

Cells have protein receptors on the extracellular surfaces of their plasma membranes. When a molecule within the matrix binds to the receptor, it changes the molecular structure of the receptor. The receptor, in turn, changes the conformation of the microfilaments positioned just inside the plasma membrane. These conformational changes induce chemical signals inside the cell that reach the nucleus and turn “on” or “off” the transcription of specific sections of DNA, which affects the production of associated proteins, thus changing the activities within the cell.

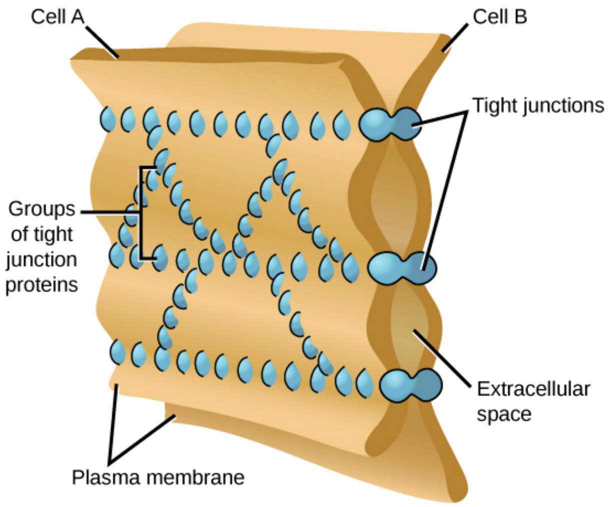
Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates the adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and initiates a series of steps that stimulate the platelets to produce clotting factors.

Intercellular Junctions

Cells can also communicate with each other via direct contact, referred to as intercellular junctions. Important animal cell contacts include tight junctions, gap junctions, and desmosomes.

Tight Junctions

A **tight junction** is a watertight seal between two adjacent animal cells ([link]). The cells are held tightly against each other by proteins (predominantly two proteins called claudins and occludins).

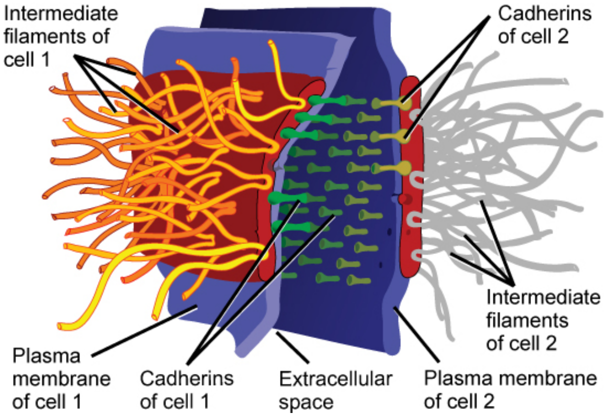


Tight junctions form watertight connections between adjacent animal cells. Proteins create tight junction adherence. (credit: modification of work by Mariana Ruiz Villareal)

This tight adherence prevents materials from leaking between the cells; tight junctions are typically found in epithelial tissues that line internal organs and cavities, and comprise most of the skin. For example, the tight junctions of the epithelial cells lining your urinary bladder prevent urine from leaking out into the extracellular space.

Desmosomes

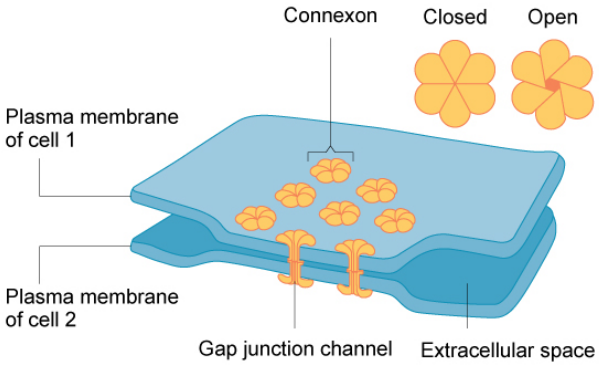
Also found only in animal cells are **desmosomes**, which act like spot welds between adjacent epithelial cells ([link]). Short proteins called cadherins in the plasma membrane connect to intermediate filaments to create desmosomes. The cadherins join two adjacent cells together and maintain the cells in a sheet-like formation in organs and tissues that stretch, like the skin, heart, and muscles.


Intermediate filaments of cell 1
Cadherins of cell 2
Intermediate filaments of cell 2
Plasma membrane of cell 1
Cadherins of cell 1
Extracellular space
Plasma membrane of cell 2

A desmosome forms a very strong spot weld between cells. It is created by the linkage of cadherins and intermediate filaments. (credit: modification of work by Mariana Ruiz Villareal)

Gap Junctions

Gap junctions provide channels between adjacent cells that allow for the transport of ions, nutrients, and other substances that enable cells to communicate ([link]).



A gap junction is a protein-lined pore that allows water and small molecules to pass between adjacent animal cells. (credit: modification of work by Mariana Ruiz Villareal)

Gap junctions develop when a set of six proteins (called connexins) in the plasma membrane arrange themselves in an elongated donut-like configuration called a connexon. When the pores (“doughnut holes”) of connexons in adjacent animal cells align, a channel between the two cells forms. Gap junctions are particularly important in cardiac muscle: The electrical signal for the muscle to contract is passed efficiently through gap junctions, allowing the heart muscle cells to contract in tandem. They also are found in single-unit smooth muscle (the most common type) found in organs like the blood vessels, intestines, and respiratory tract.

Note:
Link to Learning

To conduct a virtual microscopy lab and review the parts of a cell, work through the steps of this interactive assignment.

Section Summary

Animal cells communicate via their extracellular matrices and are connected to each other via tight junctions, desmosomes, and gap junctions. Plant cells are connected and communicate with each other via plasmodesmata.

When protein receptors on the surface of the plasma membrane of an animal cell bind to a substance in the extracellular matrix, a chain of reactions begins that changes activities taking place within the cell. A tight junction is a watertight seal between two adjacent cells, while a desmosome acts like a spot weld.

Review Questions

Exercise:

Problem:

The key components of desmosomes are cadherins and ___________.

a. actin
b. microfilaments
c. intermediate filaments
d. microtubules

Solution:

C

Free Response

Exercise:

Problem: Explain how the extracellular matrix functions.

Solution:

The extracellular matrix functions in support and attachment for animal tissues. It also functions in the healing and growth of the tissue.

Glossary

desmosome
linkages between adjacent epithelial cells that form when cadherins in the plasma membrane attach to intermediate filaments

extracellular matrix
material (primarily collagen, glycoproteins, and proteoglycans) secreted from animal cells that provides mechanical protection and anchoring for the cells in the tissue

gap junction
channel between two adjacent animal cells that allows ions, nutrients, and low molecular weight substances to pass between cells, enabling the cells to communicate

plasmodesma
(plural = plasmodesmata) channel that passes between the cell walls of adjacent plant cells, connects their cytoplasm, and allows materials to be transported from cell to cell

tight junction
firm seal between two adjacent animal cells created by protein adherence

Cell Growth and Division

By the end of this section, you will be able to:

- Describe the stages of the cell cycle
- Discuss how the cell cycle is regulated
- Describe the implications of losing control over the cell cycle
- Describe the stages of mitosis and cytokinesis, in order

So far in this chapter, you have read numerous times of the importance and prevalence of cell division. While there are a few cells in the body that do not undergo cell division (such as gametes, red blood cells, most neurons, and some muscle cells), most somatic cells divide regularly. A **somatic cell** is a general term for a body cell, and all human cells, except for the cells that produce eggs and sperm (which are referred to as germ cells), are somatic cells. Somatic cells contain *two* copies of each of their chromosomes (one copy received from each parent). A **homologous** pair of chromosomes is the two copies of a single chromosome found in each somatic cell. The human is a **diploid** organism, having 23 homologous pairs of chromosomes in each of the somatic cells. The condition of having pairs of chromosomes is known as diploidy.

Cells in the body replace themselves over the lifetime of a person. For example, the cells lining the gastrointestinal tract must be frequently replaced when constantly “worn off” by the movement of food through the gut. But what triggers a cell to divide, and how does it prepare for and complete cell division? The **cell cycle** is the sequence of events in the life of the cell from the moment it is created at the end of a previous cycle of cell division until it then divides itself, generating two new cells.

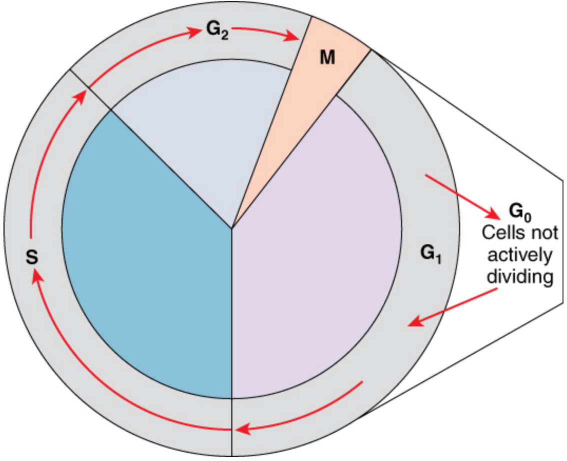
The Cell Cycle

One “turn” or cycle of the cell cycle consists of two general phases: interphase, followed by mitosis and cytokinesis. **Interphase** is the period of the cell cycle during which the cell is not dividing. The majority of cells are in interphase most of the time. **Mitosis** is the division of genetic material, during which the cell nucleus breaks down and two new, fully functional, nuclei are formed. **Cytokinesis** divides the cytoplasm into two distinctive cells.

Interphase

A cell grows and carries out all normal metabolic functions and processes in a period called G_1 ([link]). **G_1 phase** (gap 1 phase) is the first gap, or growth phase in the cell cycle. For cells that will divide again, G_1 is followed by replication of the DNA, during the S phase. The **S phase** (synthesis phase) is period during which a cell replicates its DNA.

Cell Cycle



The two major phases of the cell cycle include mitosis (cell division), and interphase, when the cell grows and performs all of its normal functions. Interphase is further subdivided into G_1, S, and G_2 phases.

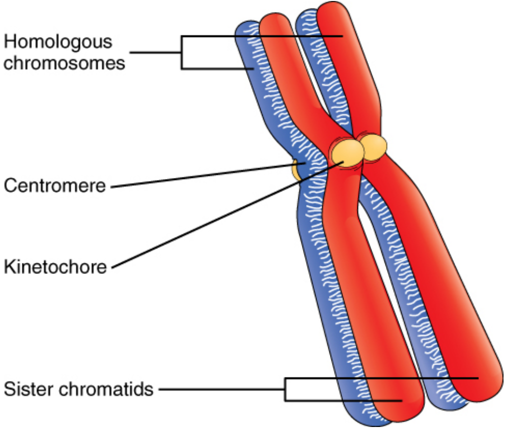
After the synthesis phase, the cell proceeds through the G_2 phase. The **G_2 phase** is a second gap phase, during which the cell continues to grow and makes the necessary preparations for mitosis. Between G_1, S, and G_2

phases, cells will vary the most in their duration of the G1 phase. It is here that a cell might spend a couple of hours, or many days. The S phase typically lasts between 8-10 hours and the G_2 phase approximately 5 hours. In contrast to these phases, the **G_0 phase** is a resting phase of the cell cycle. Cells that have temporarily stopped dividing and are resting (a common condition) and cells that have permanently ceased dividing (like nerve cells) are said to be in G_0.

The Structure of Chromosomes

Billions of cells in the human body divide every day. During the synthesis phase (S, for DNA synthesis) of interphase, the amount of DNA within the cell precisely doubles. Therefore, after DNA replication but before cell division, each cell actually contains *two* copies of each chromosome. Each copy of the chromosome is referred to as a **sister chromatid** and is physically bound to the other copy. The **centromere** is the structure that attaches one sister chromatid to another. Because a human cell has 46 chromosomes, during this phase, there are 92 chromatids (46 × 2) in the cell. Make sure not to confuse the concept of a pair of chromatids (one chromosome and its exact copy attached during mitosis) and a homologous pair of chromosomes (two paired chromosomes which were inherited separately, one from each parent) ([link]).

A Homologous Pair of Chromosomes with their Attached Sister Chromatids



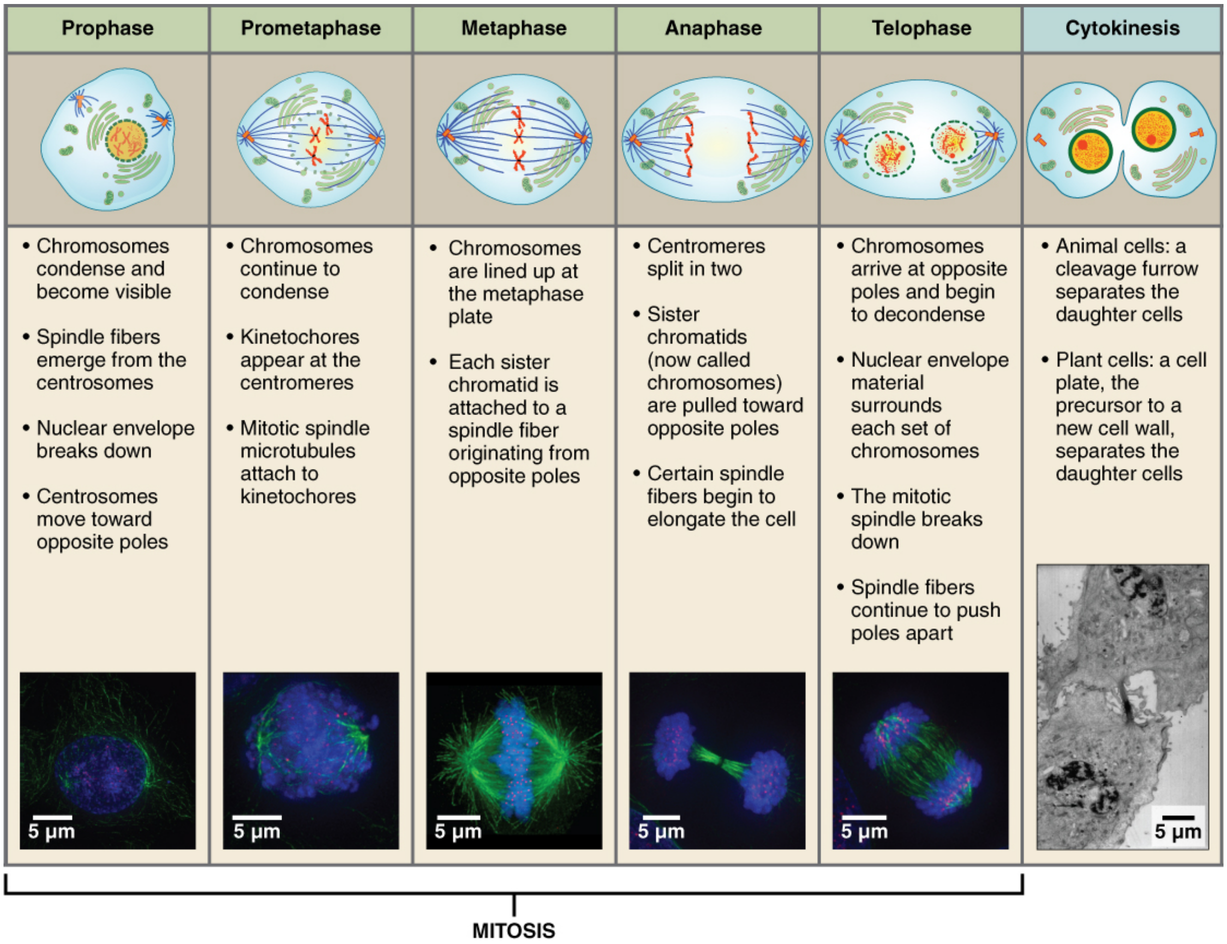
The red and blue colors correspond to a

homologous pair of chromosomes. Each member of the pair was separately inherited from one parent. Each chromosome in the homologous pair is also bound to an identical sister chromatid, which is produced by DNA replication, and results in the familiar “X” shape.

Mitosis and Cytokinesis

The **mitotic phase** of the cell typically takes between 1 and 2 hours. During this phase, a cell undergoes two major processes. First, it completes mitosis, during which the contents of the nucleus are equitably pulled apart and distributed between its two halves. Cytokinesis then occurs, dividing the cytoplasm and cell body into two new cells. Mitosis is divided into four major stages that take place after interphase ([link]) and in the following order: prophase, metaphase, anaphase, and telophase. The process is then followed by cytokinesis.

Cell Division: Mitosis Followed by Cytokinesis


Prophase
Prometaphase
Metaphase
Anaphase
Telophase
Cytokinesis
• Chromosomes condense and become visible
• Spindle fibers emerge from the centrosomes
• Nuclear envelope breaks down
• Centrosomes move toward opposite poles
• Chromosomes continue to condense
• Kinetochores appear at the centromeres
• Mitotic spindle microtubules attach to kinetochores
• Chromosomes are lined up at the metaphase plate
• Each sister chromatid is attached to a spindle fiber originating from opposite poles
• Centromeres split in two
• Sister chromatids (now called chromosomes) are pulled toward opposite poles
• Certain spindle fibers begin to elongate the cell
• Chromosomes arrive at opposite poles and begin to decondense
• Nuclear envelope material surrounds each set of chromosomes
• The mitotic spindle breaks down
• Spindle fibers continue to push poles apart
• Animal cells: a cleavage furrow separates the daughter cells
• Plant cells: a cell plate, the precursor to a new cell wall, separates the daughter cells
5 µm
5 µm
5 µm
5 µm
5 µm
5 µm
MITOSIS

The stages of cell division oversee the separation of identical genetic material into two new nuclei, followed by the division of the cytoplasm.

Prophase is the first phase of mitosis, during which the loosely packed chromatin coils and condenses into visible chromosomes. During prophase, each chromosome becomes visible with its identical partner attached, forming the familiar X-shape of sister chromatids. The nucleolus disappears early during this phase, and the nuclear envelope also disintegrates.

A major occurrence during prophase concerns a very important structure that contains the origin site for microtubule growth. Recall the cellular structures called centrioles that serve as origin points from which microtubules extend. These tiny structures also play a very important role during mitosis. A **centrosome** is a pair of centrioles together. The cell contains two centrosomes side-by-side, which begin to move apart during prophase. As the centrosomes migrate to two different sides of the cell, microtubules begin to extend from each like long fingers from two hands extending toward each other. The **mitotic spindle** is the structure composed of the centrosomes and their emerging microtubules.

Near the end of prophase there is an invasion of the nuclear area by microtubules from the mitotic spindle. The nuclear membrane has disintegrated, and the microtubules attach themselves to the centromeres that adjoin pairs of sister chromatids. The **kinetochore** is a protein structure on the centromere that is the point of attachment between the mitotic spindle and the sister chromatids. This stage is referred to as late prophase or “prometaphase” to indicate the transition between prophase and metaphase.

Metaphase is the second stage of mitosis. During this stage, the sister chromatids, with their attached microtubules, line up along a linear plane in the middle of the cell. A metaphase plate forms between the centrosomes that are now located at either end of the cell. The **metaphase plate** is the name for the plane through the center of the spindle on which the sister chromatids are positioned. The microtubules are now poised to pull apart the sister chromatids and bring one from each pair to each side of the cell.

Anaphase is the third stage of mitosis. Anaphase takes place over a few minutes, when the pairs of sister chromatids are separated from one another, forming individual chromosomes once again. These chromosomes are pulled to opposite ends of the cell by their kinetochores, as the microtubules shorten. Each end of the cell receives one partner from each pair of sister chromatids, ensuring that the two new daughter cells will contain identical genetic material.

Telophase is the final stage of mitosis. Telophase is characterized by the formation of two new daughter nuclei at either end of the dividing cell. These newly formed nuclei surround the genetic material, which uncoils such that the chromosomes return to loosely packed chromatin. Nucleoli also reappear within the new nuclei, and the mitotic spindle breaks apart, each new cell receiving its own complement of DNA, organelles, membranes, and centrioles. At this point, the cell is already beginning to split in half as cytokinesis begins.

The **cleavage furrow** is a contractile band made up of microfilaments that forms around the midline of the cell during cytokinesis. (Recall that microfilaments consist of actin.) This contractile band squeezes the two cells apart until they finally separate. Two new cells are now formed. One of these cells (the “stem cell”) enters its own cell cycle; able to grow and divide again at some future time. The other cell transforms into the functional cell of the tissue, typically replacing an “old” cell there.

Imagine a cell that completed mitosis but never underwent cytokinesis. In some cases, a cell may divide its genetic material and grow in size, but fail to undergo cytokinesis. This results in larger cells with more than one nucleus. Usually this is an unwanted aberration and can be a sign of cancerous cells.

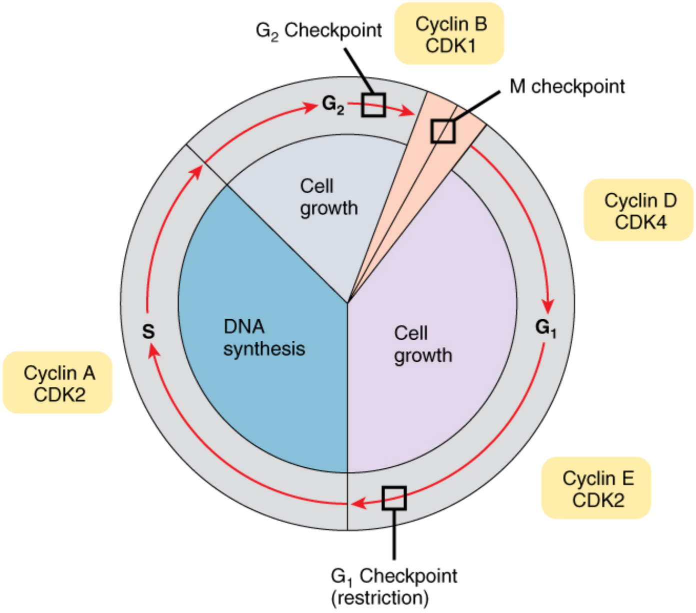
Cell Cycle Control

A very elaborate and precise system of regulation controls direct the way cells proceed from one phase to the next in the cell cycle and begin mitosis. The control system involves molecules within the cell as well as external triggers. These internal and external control triggers provide “stop” and “advance” signals for the cell. Precise regulation of the cell cycle is critical for maintaining the health of an organism, and loss of cell cycle control can lead to cancer.

Mechanisms of Cell Cycle Control

As the cell proceeds through its cycle, each phase involves certain processes that must be completed before the cell should advance to the next phase. A **checkpoint** is a point in the cell cycle at which the cycle can be signaled to move forward or stopped. At each of these checkpoints, different varieties of molecules provide the stop or go signals, depending on certain conditions within the cell. A **cyclin** is one of the primary classes of cell cycle control molecules ([link]). A **cyclin-dependent kinase (CDK)** is one of a group of molecules that work together with cyclins to determine progression past cell checkpoints. By interacting with many additional molecules, these triggers push the cell cycle forward unless prevented from doing so by “stop” signals, if for some reason the cell is not ready. At the G_1 checkpoint, the cell must be ready for DNA synthesis to occur. At the G_2 checkpoint the cell must be fully prepared for mitosis. Even during mitosis, a crucial stop and go checkpoint in metaphase ensures that the cell is fully prepared to complete cell division. The metaphase checkpoint ensures that all sister chromatids are properly attached to their respective microtubules and lined up at the metaphase plate before the signal is given to separate them during anaphase.

Control of the Cell Cycle



Cells proceed through the cell cycle under the control of a variety of molecules, such as cyclins and cyclin-dependent kinases. These control molecules determine whether or not the cell is prepared to move into the following stage.

The Cell Cycle Out of Control: Implications

Most people understand that cancer or tumors are caused by abnormal cells that multiply continuously. If the abnormal cells continue to divide unstopped, they can damage the tissues around them, spread to other parts of the body, and eventually result in death. In healthy cells, the tight regulation mechanisms of the cell cycle prevent this from happening, while failures of cell cycle control can cause unwanted and excessive cell division. Failures of control may be caused by inherited genetic abnormalities that compromise the function of certain “stop” and “go” signals. Environmental insult that damages DNA can also cause dysfunction in those signals. Often, a combination of both genetic predisposition and environmental factors lead to cancer.

The process of a cell escaping its normal control system and becoming cancerous may actually happen throughout the body quite frequently. Fortunately, certain cells of the immune system are capable of recognizing cells that have become cancerous and destroying them. However, in certain cases the cancerous cells remain undetected and continue to proliferate. If the resulting tumor does not pose a threat to surrounding tissues, it is said to be benign and can usually be easily removed. If capable of damage, the tumor is considered malignant and the patient is diagnosed with cancer.

Note:

Homeostatic Imbalances

Cancer Arises from Homeostatic Imbalances

Cancer is an extremely complex condition, capable of arising from a wide variety of genetic and environmental causes. Typically, mutations or aberrations in a cell’s DNA that compromise normal cell cycle control systems lead to cancerous tumors. Cell cycle control is an example of a homeostatic mechanism that maintains proper cell function and health. While progressing through the phases of the cell cycle, a large variety of intracellular molecules provide stop and go signals to regulate movement forward to the next phase. These signals are maintained in an intricate balance so that the cell only proceeds to the next phase when it is ready. This homeostatic control of the cell cycle can be thought of like a car’s cruise control. Cruise control will continually apply just the right amount of acceleration to maintain a desired speed, unless the driver hits the brakes, in which case the car will slow down. Similarly, the cell includes molecular messengers, such as cyclins, that push the cell forward in its cycle.

In addition to cyclins, a class of proteins that are encoded by genes called proto-oncogenes provide important signals that regulate the cell cycle and move it forward. Examples of proto-oncogene products include cell-surface receptors for growth factors, or cell-signaling molecules, two classes of molecules that can promote DNA replication and cell division.

In contrast, a second class of genes known as tumor suppressor genes sends stop signals during a cell cycle. For example, certain protein products of tumor suppressor genes signal potential problems with the DNA and thus stop the cell from dividing, while other proteins signal the cell to die if it is damaged beyond repair. Some tumor suppressor proteins also signal a sufficient surrounding cellular density, which indicates that the cell need not presently divide. The latter function is uniquely important in preventing tumor growth: normal cells exhibit a phenomenon called “contact inhibition;” thus, extensive cellular contact with neighboring cells causes a signal that stops further cell division.

These two contrasting classes of genes, proto-oncogenes and tumor suppressor genes, are like the accelerator and brake pedal of the cell’s own “cruise control system,” respectively. Under normal conditions, these stop and go signals are maintained in a homeostatic balance. Generally speaking, there are two ways that the cell’s cruise control can lose control: a malfunctioning (overactive) accelerator, or a malfunctioning (underactive) brake. When compromised through a mutation, or otherwise altered, proto-oncogenes can be converted to oncogenes, which produce oncoproteins that push a cell forward in its cycle and stimulate cell division even when it is undesirable to do so. For example, a cell that should be programmed to self-destruct (a process called apoptosis) due to extensive DNA damage might instead be triggered to proliferate by an oncoprotein. On the other hand, a dysfunctional tumor suppressor gene may fail to provide the cell with a necessary stop signal, also resulting in unwanted cell division and proliferation.

A delicate homeostatic balance between the many proto-oncogenes and tumor suppressor genes delicately controls the cell cycle and ensures that only healthy cells replicate. Therefore, a disruption of this homeostatic balance can cause aberrant cell division and cancerous growths.

Note:

Visit this link to learn about mitosis. Mitosis results in two identical diploid cells. What structures forms during prophase?

Chapter Review

The life of cell consists of stages that make up the cell cycle. After a cell is born, it passes through an interphase before it is ready to replicate itself and produce daughter cells. This interphase includes two gap phases (G_1 and G_2), as well as an S phase, during which its DNA is replicated in preparation for cell division. The cell cycle is under precise regulation by chemical messengers both inside and outside the cell that provide “stop” and “go” signals for movement from one phase to the next. Failures of these signals can result in cells that continue to divide uncontrollably, which can lead to cancer.

Once a cell has completed interphase and is ready for cell division, it proceeds through four separate stages of mitosis (prophase, metaphase, anaphase, and telophase). Telophase is followed by the division of the cytoplasm (cytokinesis), which generates two daughter cells. This process takes place in all normally dividing cells of the body except for the germ cells that produce eggs and sperm.

Interactive Link Questions

Exercise:

Problem:

Visit this link to learn about mitosis. Mitosis results in two identical diploid cells. What structures form during prophase?

Solution:

the spindle

Review Questions

Exercise:

Problem:

Which of the following phases is characterized by preparation for DNA synthesis?

a. G_0
b. G_1
c. G_2
d. S

Solution:

B

Exercise:

Problem:

A mutation in the gene for a cyclin protein might result in which of the following?

a. a cell with additional genetic material than normal

b. cancer
c. a cell with less genetic material than normal
d. any of the above

Solution:

D

Exercise:

Problem: What is a primary function of tumor suppressor genes?

a. stop all cells from dividing
b. stop certain cells from dividing
c. help oncogenes produce oncoproteins
d. allow the cell to skip certain phases of the cell cycle

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

What would happen if anaphase proceeded even though the sister chromatids were not properly attached to their respective microtubules and lined up at the metaphase plate?

Solution:

One or both of the new daughter cells would accidently receive duplicate chromosomes and/or would be missing certain chromosomes.

Exercise:

Problem:

What are cyclins and cyclin-dependent kinases, and how do they interact?

Solution:

A cyclin is one of the primary classes of cell cycle control molecules, while a cyclin-dependent kinase (is one of a group of molecules that work together with cyclins to determine progression past cell checkpoints. By interacting with many additional molecules, these triggers push the cell cycle forward unless prevented from doing so by “stop” signals, if for some reason the cell is not ready.

Glossary

anaphase
third stage of mitosis (and meiosis), during which sister chromatids separate into two new nuclear regions of a dividing cell

cell cycle
life cycle of a single cell, from its birth until its division into two new daughter cells

centromere
region of attachment for two sister chromatids

centrosome
cellular structure that organizes microtubules during cell division

checkpoint
progress point in the cell cycle during which certain conditions must be met in order for the cell to proceed to a subsequence phase

cleavage furrow

contractile ring that forms around a cell during cytokinesis that pinches the cell into two halves

cyclin
one of a group of proteins that function in the progression of the cell cycle

cyclin-dependent kinase (CDK)
one of a group of enzymes associated with cyclins that help them perform their functions

cytokinesis
final stage in cell division, where the cytoplasm divides to form two separate daughter cells

diploid
condition marked by the presence of a double complement of genetic material (two sets of chromosomes, one set inherited from each of two parents)

G_0 phase
phase of the cell cycle, usually entered from the G_1 phase; characterized by long or permanent periods where the cell does not move forward into the DNA synthesis phase

G_1 phase
first phase of the cell cycle, after a new cell is born

G_2 phase
third phase of the cell cycle, after the DNA synthesis phase

homologous
describes two copies of the same chromosome (not identical), one inherited from each parent

interphase
entire life cycle of a cell, excluding mitosis

kinetochore
region of a centromere where microtubules attach to a pair of sister chromatids

metaphase
second stage of mitosis (and meiosis), characterized by the linear alignment of sister chromatids in the center of the cell

metaphase plate
linear alignment of sister chromatids in the center of the cell, which takes place during metaphase

mitosis
division of genetic material, during which the cell nucleus breaks down and two new, fully functional, nuclei are formed

mitotic phase
phase of the cell cycle in which a cell undergoes mitosis

mitotic spindle
network of microtubules, originating from centrioles, that arranges and pulls apart chromosomes during mitosis

prophase
first stage of mitosis (and meiosis), characterized by breakdown of the nuclear envelope and condensing of the chromatin to form chromosomes

S phase
stage of the cell cycle during which DNA replication occurs

sister chromatid
one of a pair of identical chromosomes, formed during DNA replication

somatic cell
all cells of the body excluding gamete cells

telophase

final stage of mitosis (and meiosis), preceding cytokinesis, characterized by the formation of two new daughter nuclei

The Special Case of Sex Cells

By the end of this section, you will be able to:

- Describe the behavior of chromosomes during meiosis
- Describe cellular events during meiosis
- Explain the differences between meiosis and mitosis
- Explain the mechanisms within meiosis that generate genetic variation among the products of meiosis

Sexual reproduction requires **fertilization**, the union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. Haploid cells contain one set of chromosomes. Cells containing two sets of chromosomes are called diploid. The number of sets of chromosomes in a cell is called its ploidy level. If the reproductive cycle is to continue, then the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division that reduces the number of chromosome sets.

Most animals and plants are diploid, containing two sets of chromosomes. In each **somatic cell** of the organism (all cells of a multicellular organism except the gametes or reproductive cells), the nucleus contains two copies of each chromosome, called homologous chromosomes. Somatic cells are sometimes referred to as “body” cells. Homologous chromosomes are matched pairs containing the same genes in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent; all together, they are considered a full set of chromosomes. Haploid cells, containing a single copy of each homologous chromosome, are found only within structures that give rise to either gametes or spores. **Spores** are haploid cells that can produce a haploid organism or can fuse with another spore to form a diploid cell. All animals and most plants produce eggs and sperm, or gametes. Some plants and all fungi produce spores.

The nuclear division that forms haploid cells, which is called **meiosis**, is related to mitosis. As you have learned, mitosis is the part of a cell reproduction cycle that results in identical daughter nuclei that are also genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei are at the same ploidy level—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve this reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the major process and the stages are designated with a “I” or a “II.” Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. **Meiosis II**, in which the second round of meiotic division takes place, includes prophase II, prometaphase II, and so on.

Meiosis I

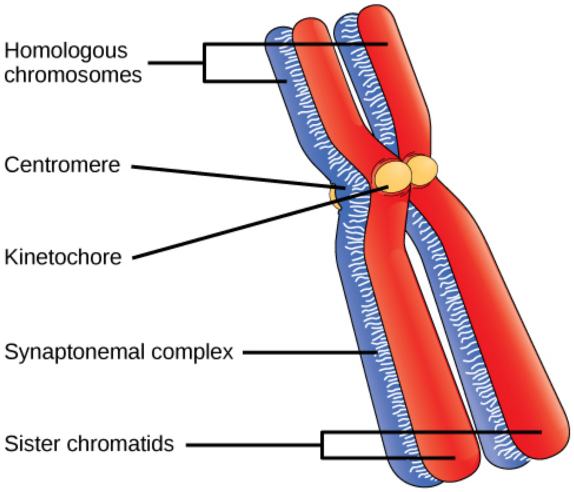
Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase, which is also called the first gap phase, is the first phase of the interphase and is focused on cell growth. The S phase is the second phase of interphase, during which the DNA of the chromosomes is replicated. Finally, the G_2 phase, also called the second gap phase, is the third and final phase of interphase; in this phase, the cell undergoes the final preparations for meiosis.

During DNA duplication in the S phase, each chromosome is replicated to produce two identical copies, called sister chromatids, that are held together at the centromere by **cohesin** proteins. Cohesin holds the chromatids together until anaphase II. The centrosomes, which are the structures that organize the microtubules of the meiotic spindle, also replicate. This prepares the cell to enter prophase I, the first meiotic phase.

Prophase I

Early in prophase I, before the chromosomes can be seen clearly microscopically, the homologous chromosomes are attached at their tips to the nuclear envelope by proteins. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. Recall that, in mitosis, homologous chromosomes do not pair together. In mitosis, homologous chromosomes line up end-to-end so that when they divide, each daughter cell receives a sister chromatid from both members of the homologous pair. The **synaptonemal complex**, a lattice of proteins between the homologous chromosomes, first forms at specific locations and then spreads to cover the entire length of the chromosomes. The tight pairing of the homologous chromosomes is called **synapsis**. In synapsis, the genes on the chromatids of the homologous chromosomes are aligned precisely with each other. The synaptonemal complex supports the exchange of chromosomal segments between non-sister homologous chromatids, a process called crossing over. Crossing over can be observed visually after the exchange as **chiasmata** (singular = chiasma) ([link]).

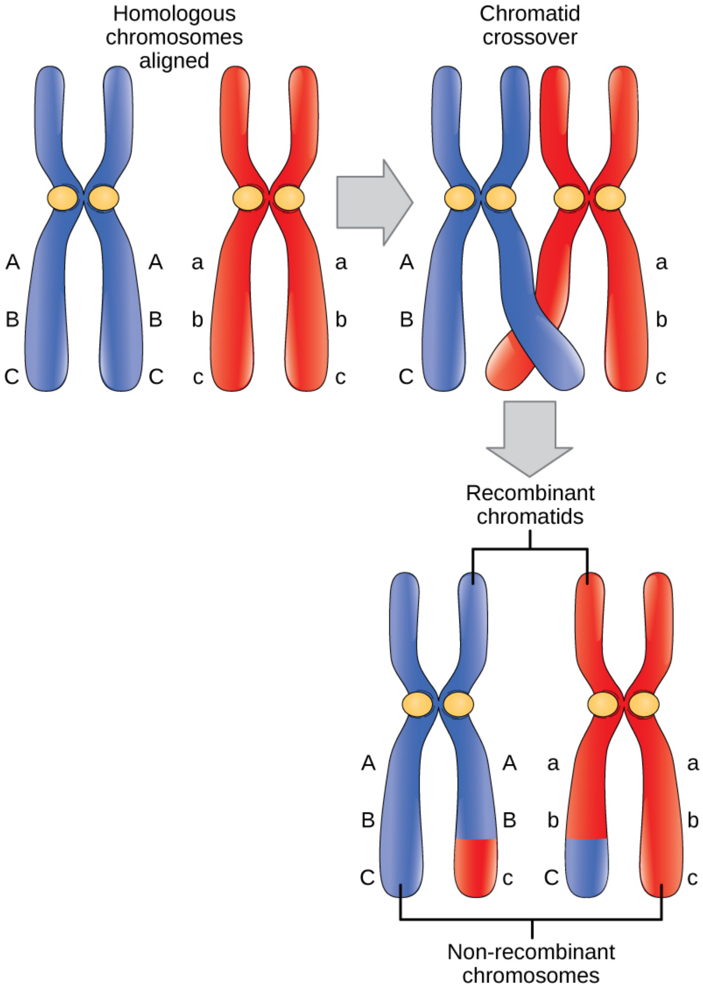
In species such as humans, even though the X and Y sex chromosomes are not homologous (most of their genes differ), they have a small region of homology that allows the X and Y chromosomes to pair up during prophase I. A partial synaptonemal complex develops only between the regions of homology.



Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

Located at intervals along the synaptonemal complex are large protein assemblies called **recombination nodules**. These assemblies mark the points of later chiasmata and mediate the multistep process of **crossover**—or genetic recombination—between the non-sister chromatids. Near the recombination nodule on each chromatid, the double-stranded DNA is cleaved, the cut ends are modified, and a new connection is made between the non-sister chromatids. As prophase I progresses, the synaptonemal complex begins to break down and the chromosomes begin to condense. When the synaptonemal complex is gone, the homologous chromosomes remain attached to each other at the centromere and at chiasmata. The chiasmata remain until anaphase I. The number of chiasmata varies according to the species and the length of the chromosome. There must be at least one chiasma per chromosome for proper separation of homologous chromosomes during meiosis I, but there may be as many as 25. Following crossover, the synaptonemal complex breaks down and the cohesin connection between homologous pairs is also removed. At the end of prophase I, the pairs are held together only at the chiasmata ([link]) and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation in the nuclei produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete cell it will carry some DNA from one parent of the individual and some DNA from the other parent. The sister recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover. Multiple crossovers in an arm of the chromosome have the same effect, exchanging segments of DNA to create recombinant chromosomes.



Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes.

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. Kinetochore proteins are multiprotein complexes that bind the centromeres of a chromosome to the microtubules of the mitotic spindle. Microtubules grow from centrosomes placed at opposite poles of the cell. The microtubules move toward the middle of the cell and attach to one of the two fused homologous chromosomes. The microtubules attach at each chromosomes' kinetochores. With each member of the homologous pair attached to opposite poles of the cell, in the next phase, the microtubules can pull the homologous pair apart. A spindle fiber that has attached to a kinetochore is called a kinetochore microtubule. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome facing each pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

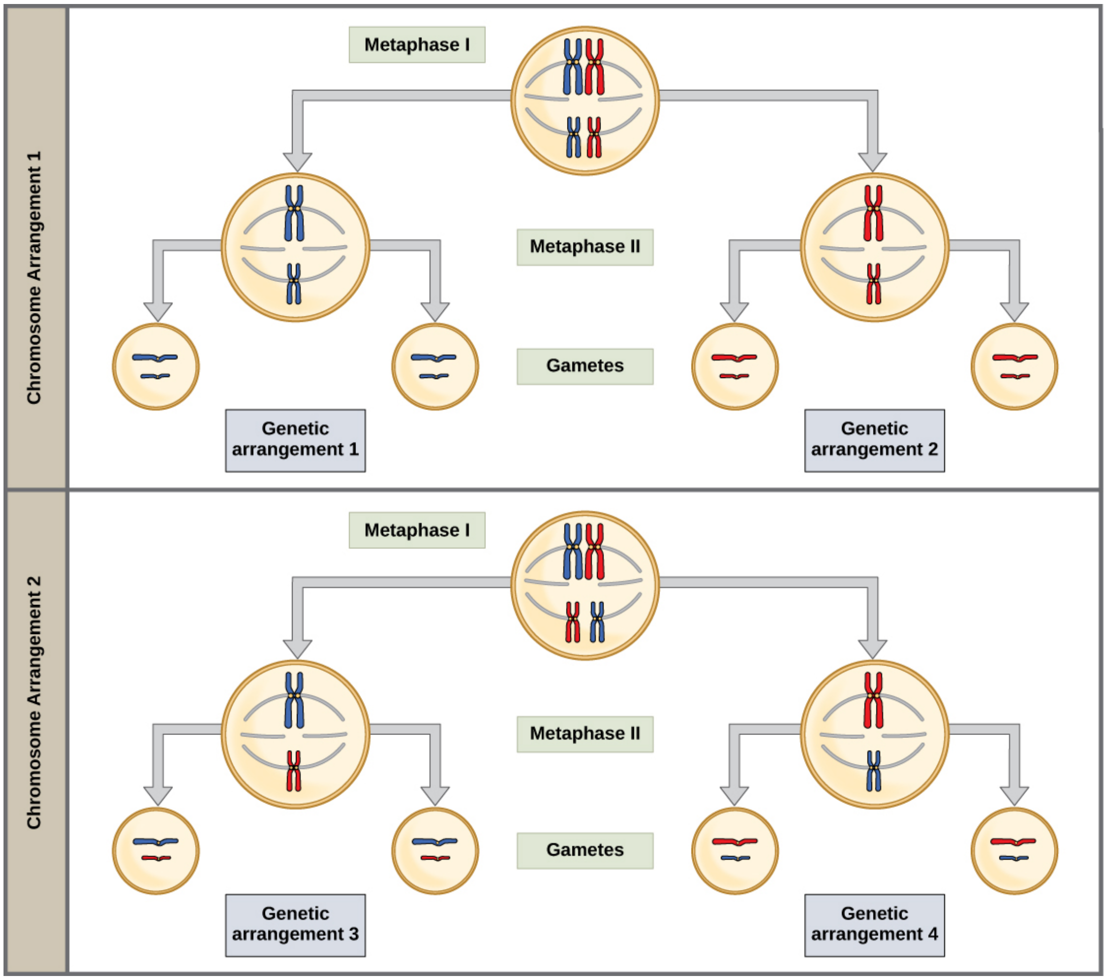
Metaphase I

During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The homologous pairs orient themselves randomly at the equator. For example, if the two homologous members of chromosome 1 are labeled a and b, then the chromosomes could line up a-b, or b-a. This is important in determining the genes carried by a gamete, as each will only receive one of the two homologous chromosomes. Recall that homologous chromosomes are not identical. They contain slight differences in their genetic information, causing each gamete to have a unique genetic makeup.

This randomness is the physical basis for the creation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. Every cell of the multicellular offspring has copies of the original two sets of homologous chromosomes. In prophase I of meiosis, the homologous chromosomes form the tetrads. In metaphase I, these pairs line up at the midway point between the two poles of the cell to form the metaphase plate. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

This event—the random (or independent) assortment of homologous chromosomes at the metaphase plate—is the second mechanism that introduces variation into the gametes or spores. In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations is dependent on the number of chromosomes making up a set. There are two possibilities for orientation at the metaphase plate; the possible number of alignments therefore equals $2n$, where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possible genetically-distinct gametes. This number does not include the variability that was previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition ([link]).

To summarize the genetic consequences of meiosis I, the maternal and paternal genes are recombined by crossover events that occur between each homologous pair during prophase I. In addition, the random assortment of tetrads on the metaphase plate produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.



Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes (*n* = 2). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is 2*n*, where *n* equals the number of chromosomes in a set. In this example, there are four possible genetic combinations for the gametes. With *n* = 23 in human cells, there are over 8 million possible combinations of paternal and maternal chromosomes.

Anaphase I

In anaphase I, the microtubules pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. The chiasmata are broken in anaphase I as the microtubules attached to the fused kinetochores pull the homologous chromosomes apart ([link]).

Telophase I and Cytokinesis

In telophase, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur, depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the chromatids in telophase I. In other organisms, cytokinesis—the physical separation of the cytoplasmic components into two daughter cells—occurs without reformation of the nuclei. In nearly all species of animals and some fungi, cytokinesis separates the cell contents via a cleavage furrow (constriction of the actin ring that leads to cytoplasmic division). In plants, a cell plate is formed during cell cytokinesis by Golgi vesicles fusing at the metaphase plate. This cell plate will ultimately lead to the formation of cell walls that separate the two daughter cells.

Two haploid cells are the end result of the first meiotic division. The cells are haploid because at each pole, there is just one of each pair of the homologous chromosomes. Therefore, only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though each homolog still consists of two sister chromatids. Recall that sister chromatids are merely duplicates of one of the two homologous chromosomes (except for changes that occurred during crossing over). In meiosis II, these two sister chromatids will separate, creating four haploid daughter cells.

Note:
Link to Learning

Review the process of meiosis, observing how chromosomes align and migrate, at Meiosis: An Interactive Animation.

Meiosis II

In some species, cells enter a brief interphase, or **interkinesis**, before entering meiosis II. Interkinesis lacks an S phase, so chromosomes are not duplicated. The two cells produced in meiosis I go through the events of meiosis II in synchrony. During meiosis II, the sister chromatids within the two daughter cells separate, forming four new haploid gametes. The mechanics of meiosis II is similar to mitosis, except that each dividing cell has only one set of homologous chromosomes. Therefore, each cell has half the number of sister chromatids to separate out as a diploid cell undergoing mitosis.

Prophase II

If the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes that were duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed.

Prometaphase II

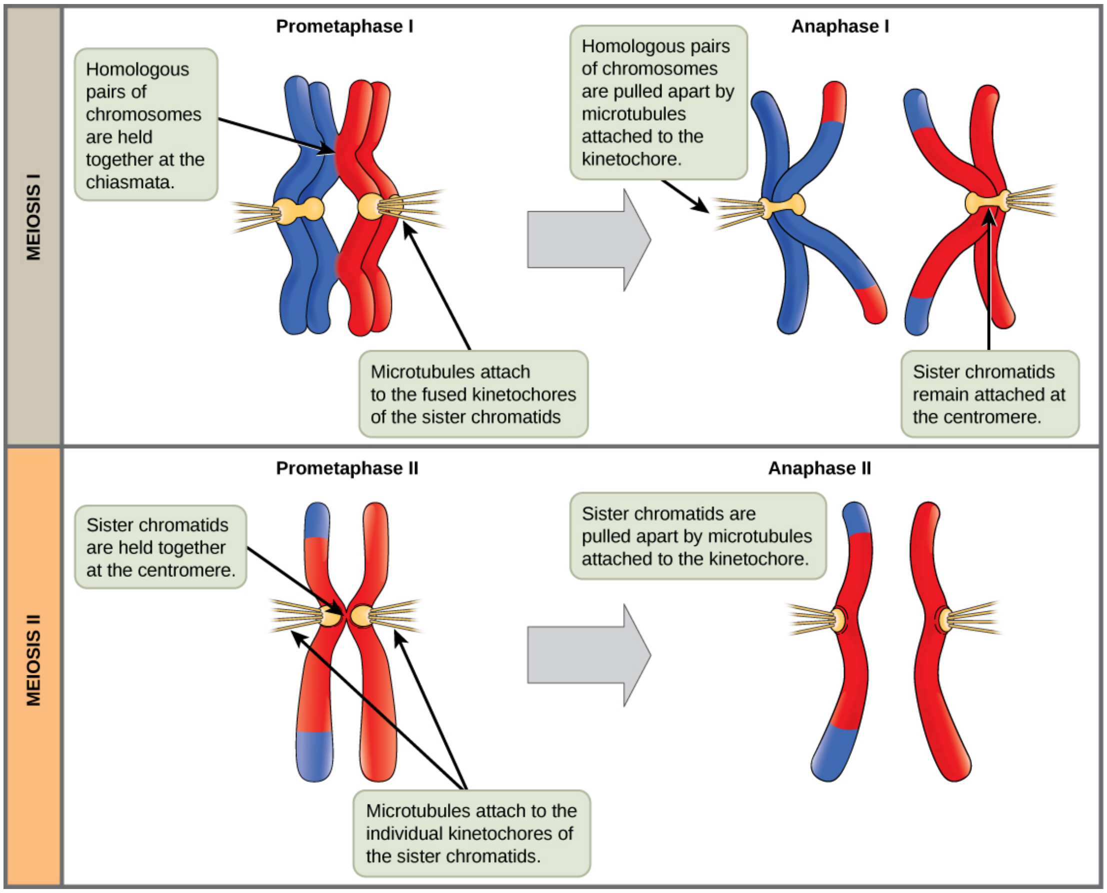
The nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles.

Metaphase II

The sister chromatids are maximally condensed and aligned at the equator of the cell.

Anaphase II

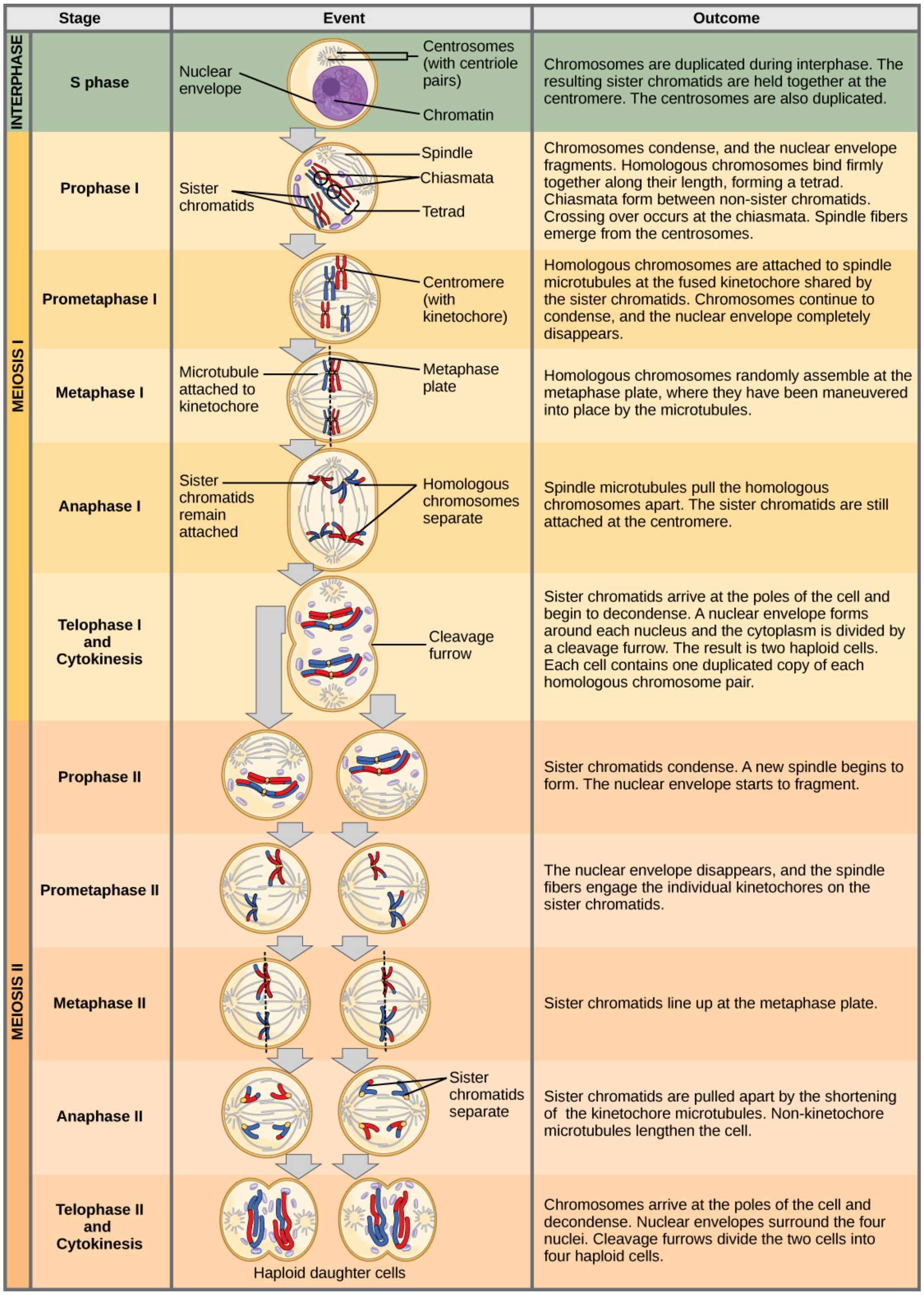
The sister chromatids are pulled apart by the kinetochore microtubules and move toward opposite poles. Non-kinetochore microtubules elongate the cell.


MEIOSIS I
Prometaphase I
Homologous pairs of chromosomes are held together at the chiasmata.
Microtubules attach to the fused kinetochores of the sister chromatids
Anaphase I
Homologous pairs of chromosomes are pulled apart by microtubules attached to the kinetochore.
Sister chromatids remain attached at the centromere.
MEIOSIS II
Prometaphase II
Sister chromatids are held together at the centromere.
Microtubules attach to the individual kinetochores of the sister chromatids.
Anaphase II
Sister chromatids are pulled apart by microtubules attached to the kinetochore.

The process of chromosome alignment differs between meiosis I and meiosis II. In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes, and the homologous chromosomes are arranged at the midpoint of the cell in metaphase I. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to the kinetochores of sister chromatids, and the sister chromatids are arranged at the midpoint of the cells in metaphase II. In anaphase II, the sister chromatids are separated.

Telophase II and Cytokinesis

The chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four unique haploid cells. At this point, the newly formed nuclei are both haploid. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombining of maternal and paternal segments of chromosomes (with their sets of genes) that occurs during crossover. The entire process of meiosis is outlined in [link].



	Stage	Event	Outcome
INTERPHASE	**S phase**	Nuclear envelope Centrosomes (with centriole pairs) Chromatin	Chromosomes are duplicated during interphase. The resulting sister chromatids are held together at the centromere. The centrosomes are also duplicated.
MEIOSIS I	**Prophase I**	Sister chromatids Spindle Chiasmata Tetrad	Chromosomes condense, and the nuclear envelope fragments. Homologous chromosomes bind firmly together along their length, forming a tetrad. Chiasmata form between non-sister chromatids. Crossing over occurs at the chiasmata. Spindle fibers emerge from the centrosomes.
	Prometaphase I	Centromere (with kinetochore)	Homologous chromosomes are attached to spindle microtubules at the fused kinetochore shared by the sister chromatids. Chromosomes continue to condense, and the nuclear envelope completely disappears.
	Metaphase I	Microtubule attached to kinetochore Metaphase plate	Homologous chromosomes randomly assemble at the metaphase plate, where they have been maneuvered into place by the microtubules.
	Anaphase I	Sister chromatids remain attached Homologous chromosomes separate	Spindle microtubules pull the homologous chromosomes apart. The sister chromatids are still attached at the centromere.
	Telophase I and Cytokinesis	Cleavage furrow	Sister chromatids arrive at the poles of the cell and begin to decondense. A nuclear envelope forms around each nucleus and the cytoplasm is divided by a cleavage furrow. The result is two haploid cells. Each cell contains one duplicated copy of each homologous chromosome pair.
MEIOSIS II	**Prophase II**		Sister chromatids condense. A new spindle begins to form. The nuclear envelope starts to fragment.
	Prometaphase II		The nuclear envelope disappears, and the spindle fibers engage the individual kinetochores on the sister chromatids.
	Metaphase II		Sister chromatids line up at the metaphase plate.
	Anaphase II	Sister chromatids separate	Sister chromatids are pulled apart by the shortening of the kinetochore microtubules. Non-kinetochore microtubules lengthen the cell.
	Telophase II and Cytokinesis	Haploid daughter cells	Chromosomes arrive at the poles of the cell and decondense. Nuclear envelopes surround the four nuclei. Cleavage furrows divide the two cells into four haploid cells.

An animal cell with a diploid number of four ($2n = 4$) proceeds through the stages of meiosis to form four haploid daughter cells.

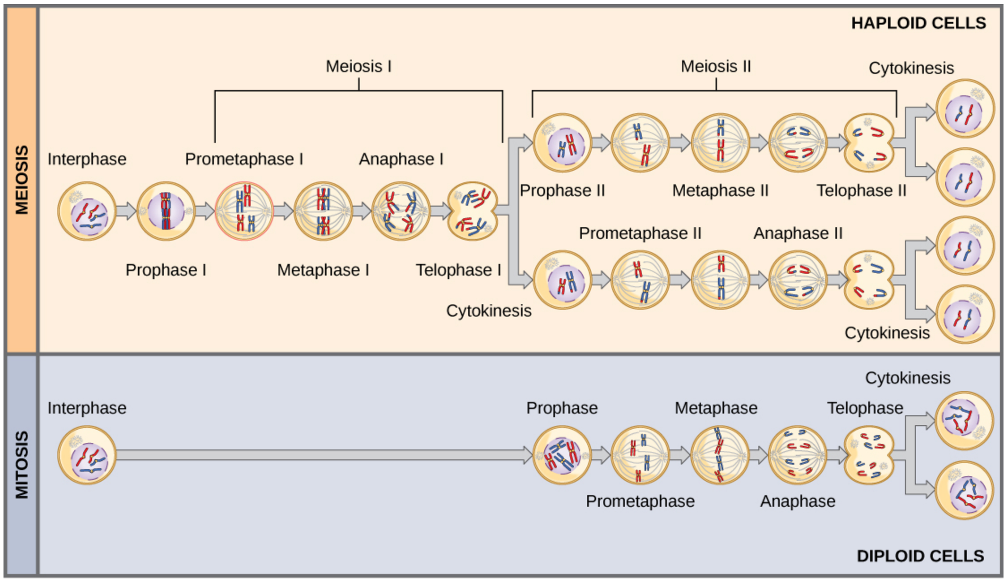
Comparing Meiosis and Mitosis

Mitosis and meiosis are both forms of division of the nucleus in eukaryotic cells. They share some similarities, but also exhibit distinct differences that lead to very different outcomes ([link]). Mitosis is a single nuclear division that results in two nuclei that are usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original nucleus. They have the same number of sets of chromosomes, one set in the case of haploid cells and two sets in the case of diploid cells. In most plants and all animal species, it is typically diploid cells that undergo mitosis to form new diploid cells. In contrast, meiosis consists of two nuclear divisions resulting in four nuclei that are usually partitioned into four new cells. The nuclei resulting from meiosis are not genetically identical and they contain one chromosome set only. This is half the number of chromosome sets in the original cell, which is diploid.

The main differences between mitosis and meiosis occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together with the synaptonemal complex, develop chiasmata and undergo crossover between sister chromatids, and line up along the metaphase plate in tetrads with kinetochore fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I.

When the chiasmata resolve and the tetrad is broken up with the homologs moving to one pole or another, the ploidy level—the number of sets of chromosomes in each future nucleus—has been reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level during mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, the duplicated chromosomes (only one set of them) line up on the metaphase plate with divided kinetochores attached to kinetochore fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid—now referred to as a chromosome—is pulled to one pole while the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossover, the two products of each individual meiosis II division would be identical (like in mitosis). Instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

Note:

Evolution Connection

The Mystery of the Evolution of Meiosis

Some characteristics of organisms are so widespread and fundamental that it is sometimes difficult to remember that they evolved like other simpler traits. Meiosis is such an extraordinarily complex series of cellular events that biologists have had trouble hypothesizing and testing how it may have evolved. Although meiosis is inextricably entwined with sexual reproduction and its advantages and disadvantages, it is important to separate the questions of the evolution of meiosis and the evolution of sex, because early meiosis may have been advantageous for different reasons than it is now. Thinking outside the box and imagining what the early benefits from meiosis might have been is one approach to uncovering how it may have evolved.

Meiosis and mitosis share obvious cellular processes and it makes sense that meiosis evolved from mitosis. The difficulty lies in the clear differences between meiosis I and mitosis. Adam Wilkins and Robin Holliday[footnote] summarized the unique events that needed to occur for the evolution of meiosis from mitosis. These steps are homologous chromosome pairing, crossover exchanges, sister chromatids remaining attached during anaphase, and suppression of DNA replication in interphase. They argue that the first step is the hardest and most important, and that understanding how it evolved would make the evolutionary process clearer. They suggest genetic experiments that might shed light on the evolution of synapsis.

Adam S. Wilkins and Robin Holliday, “The Evolution of Meiosis from Mitosis,” *Genetics* 181 (2009): 3–12.

There are other approaches to understanding the evolution of meiosis in progress. Different forms of meiosis exist in single-celled protists. Some appear to be simpler or more “primitive” forms of meiosis. Comparing the meiotic divisions of different protists may shed light on the evolution of meiosis. Marilee Ramesh and colleagues[footnote] compared the genes involved in meiosis in protists to understand when and where meiosis might have evolved. Although research is still ongoing, recent scholarship into meiosis in protists suggests that some aspects of meiosis may have evolved later than others. This kind of genetic comparison can tell us what aspects of meiosis are the oldest and what cellular processes they may have borrowed from in earlier cells.

Marilee A. Ramesh, Shehre-Banoo Malik and John M. Logsdon, Jr, “A Phylogenetic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis,” *Current Biology* 15 (2005):185–91.

Note:

Link to Learning

Click through the steps of this interactive animation to compare the meiotic process of cell division to that of mitosis: How Cells Divide.

Section Summary

Sexual reproduction requires that diploid organisms produce haploid cells that can fuse during fertilization to form diploid offspring. As with mitosis, DNA replication occurs prior to meiosis during the S-phase of the cell cycle. Meiosis is a series of events that arrange and separate chromosomes and chromatids into daughter cells. During the interphases of meiosis, each chromosome is duplicated. In meiosis, there are two rounds of nuclear division resulting in four nuclei and usually four daughter cells, each with half the number of chromosomes as the parent cell. The first separates homologs, and the second—like mitosis—separates chromatids into individual chromosomes. During meiosis, variation in the daughter nuclei is introduced because of crossover in prophase I and random alignment of tetrads at metaphase I. The cells that are produced by meiosis are genetically unique.

Meiosis and mitosis share similarities, but have distinct outcomes. Mitotic divisions are single nuclear divisions that produce daughter nuclei that are genetically identical and have the same number of chromosome sets as the original cell. Meiotic divisions include two nuclear divisions that produce four daughter nuclei that are genetically different and have one chromosome set instead of the two sets of chromosomes in the parent cell. The main differences between the processes occur in the first division of meiosis, in which homologous chromosomes are paired and exchange non-sister chromatid segments. The homologous chromosomes separate into different nuclei during meiosis I, causing a reduction of ploidy level in the first division. The second division of meiosis is more similar to a mitotic division, except that the daughter cells do not contain identical genomes because of crossover.

Review Questions

Exercise:

Problem: Meiosis produces ________ daughter cells.

a. two haploid
b. two diploid
c. four haploid
d. four diploid

Solution:

C

Exercise:

Problem: What structure is most important in forming the tetrads?

a. centromere
b. synaptonemal complex
c. chiasma
d. kinetochore

Solution:

B

Exercise:

Problem:

At which stage of meiosis are sister chromatids separated from each other?

a. prophase I
b. prophase II
c. anaphase I
d. anaphase II

Solution:

D

Exercise:

Problem:

At metaphase I, homologous chromosomes are connected only at what structures?

a. chiasmata
b. recombination nodules
c. microtubules
d. kinetochores

Solution:

A

Exercise:

Problem: Which of the following is *not* true in regard to crossover?

a. Spindle microtubules guide the transfer of DNA across the synaptonemal complex.
b. Non-sister chromatids exchange genetic material.
c. Chiasmata are formed.
d. Recombination nodules mark the crossover point.

Solution:

C

Exercise:

Problem:

What phase of mitotic interphase is missing from meiotic interkinesis?

a. G_0 phase
b. G_1 phase
c. S phase
d. G_2 phase

Solution:

C

Exercise:

Problem: The part of meiosis that is similar to mitosis is ________.

a. meiosis I
b. anaphase I
c. meiosis II

d. interkinesis

Solution:

C

Exercise:

Problem:

If a muscle cell of a typical organism has 32 chromosomes, how many chromosomes will be in a gamete of that same organism?

a. 8
b. 16
c. 32
d. 64

Solution:

B

Free Response

Exercise:

Problem: Describe the process that results in the formation of a tetrad.

Solution:

During the meiotic interphase, each chromosome is duplicated. The sister chromatids that are formed during synthesis are held together at the centromere region by cohesin proteins. All chromosomes are attached to the nuclear envelope by their tips. As the cell enters prophase I, the nuclear envelope begins to fragment, and the proteins holding homologous chromosomes locate each other. The four sister chromatids align lengthwise, and a protein lattice called the synaptonemal complex is formed between them to bind them together. The synaptonemal complex facilitates crossover between non-sister chromatids, which is observed as chiasmata along the length of the chromosome. As prophase I progresses, the synaptonemal complex breaks down and the sister chromatids become free, except where they are attached by chiasmata. At this stage, the four chromatids are visible in each homologous pairing and are called a tetrad.

Exercise:

Problem:

Explain how the random alignment of homologous chromosomes during metaphase I contributes to the variation in gametes produced by meiosis.

Solution:

Random alignment leads to new combinations of traits. The chromosomes that were originally inherited by the gamete-producing individual came equally from the egg and the sperm. In metaphase I, the duplicated copies of these maternal and paternal homologous chromosomes line up across the center of the cell. The orientation of each tetrad is random. There is an equal chance that the maternally derived chromosomes will be facing either pole. The same is true of the paternally derived chromosomes. The alignment should occur differently in almost every meiosis. As the homologous chromosomes are pulled apart in anaphase I, any combination of maternal and paternal chromosomes will move toward each pole. The gametes formed from these two groups of chromosomes will have a mixture of traits from the individual’s parents. Each gamete is unique.

Exercise:

Problem:

What is the function of the fused kinetochore found on sister chromatids in prometaphase I?

Solution:

In metaphase I, the homologous chromosomes line up at the metaphase plate. In anaphase I, the homologous chromosomes are pulled apart and move to opposite poles. Sister chromatids are not separated until meiosis II. The fused kinetochore formed during meiosis I ensures that each spindle microtubule that binds to the tetrad will attach to both sister chromatids.

Exercise:

Problem:

In a comparison of the stages of meiosis to the stages of mitosis, which stages are unique to meiosis and which stages have the same events in both meiosis and mitosis?

Solution:

All of the stages of meiosis I, except possibly telophase I, are unique because homologous chromosomes are separated, not sister chromatids. In some species, the chromosomes do not decondense and the nuclear envelopes do not form in telophase I. All of the stages of meiosis II have the same events as the stages of mitosis, with the possible exception of prophase II. In some species, the chromosomes are still condensed and there is no nuclear envelope. Other than this, all processes are the same.

Glossary

chiasmata	(singular, *chiasma*) the structure that forms at the crossover points after genetic material is exchanged
cohesin	proteins that form a complex that seals sister chromatids together at their centromeres until anaphase II of meiosis

crossover
exchange of genetic material between non-sister chromatids resulting in chromosomes that incorporate genes from both parents of the organism

fertilization
union of two haploid cells from two individual organisms

interkinesis
(also, *interphase II*) brief period of rest between meiosis I and meiosis II

meiosis
a nuclear division process that results in four haploid cells

meiosis I
first round of meiotic cell division; referred to as reduction division because the ploidy level is reduced from diploid to haploid

meiosis II
second round of meiotic cell division following meiosis I; sister chromatids are separated into individual chromosomes, and the result is four unique haploid cells

recombination nodules
protein assemblies formed on the synaptonemal complex that mark the points of crossover events and mediate the multistep process of genetic recombination between non-sister chromatids

reduction division
nuclear division that produces daughter nuclei each having one-half as many chromosome sets as the parental nucleus; meiosis I is a reduction division

somatic cell
all the cells of a multicellular organism except the gametes or reproductive cells

spore

haploid cell that can produce a haploid multicellular organism or can fuse with another spore to form a diploid cell

synapsis

formation of a close association between homologous chromosomes during prophase I

synaptonemal complex

protein lattice that forms between homologous chromosomes during prophase I, supporting crossover

tetrad

two duplicated homologous chromosomes (four chromatids) bound together by chiasmata during prophase I

Cellular Differentiation

By the end of this section, you will be able to:

- Discuss how the generalized cells of a developing embryo or the stem cells of an adult organism become differentiated into specialized cells
- Distinguish between the categories of stem cells

How does a complex organism such as a human develop from a single cell —a fertilized egg—into the vast array of cell types such as nerve cells, muscle cells, and epithelial cells that characterize the adult? Throughout development and adulthood, the process of cellular differentiation leads cells to assume their final morphology and physiology. Differentiation is the process by which unspecialized cells become specialized to carry out distinct functions.

Stem Cells

A **stem cell** is an unspecialized cell that can divide without limit as needed and can, under specific conditions, differentiate into specialized cells. Stem cells are divided into several categories according to their potential to differentiate.

The first embryonic cells that arise from the division of the zygote are the ultimate stem cells; these stems cells are described as **totipotent** because they have the potential to differentiate into any of the cells needed to enable an organism to grow and develop.

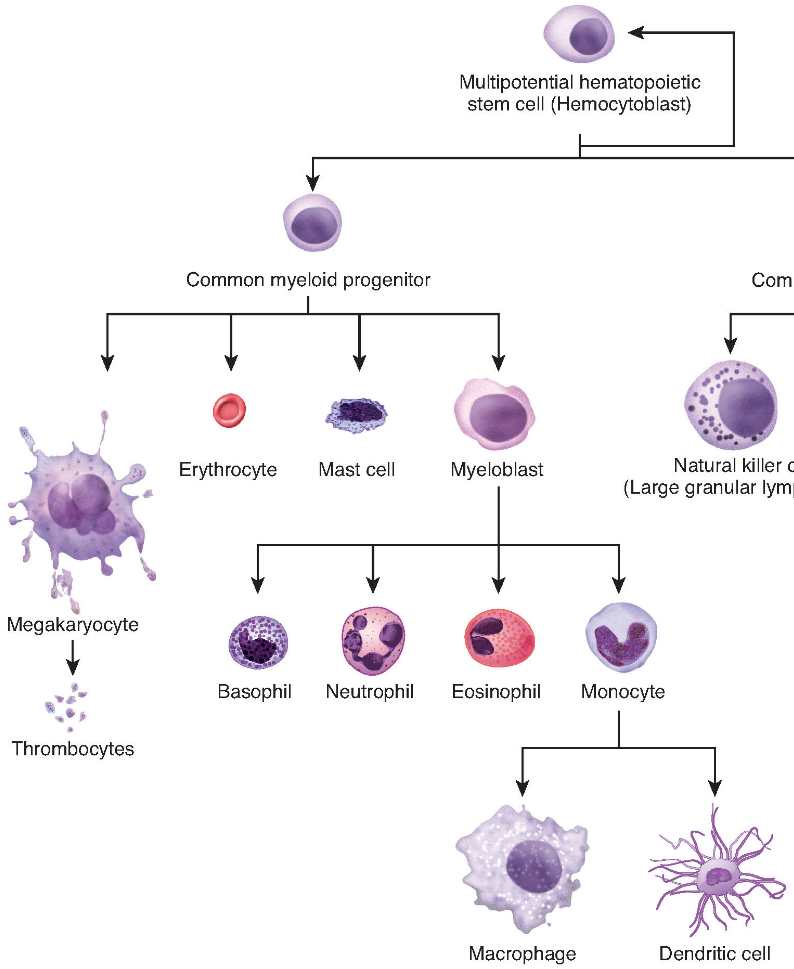
The embryonic cells that develop from totipotent stem cells and are precursors to the fundamental tissue layers of the embryo are classified as pluripotent. A **pluripotent** stem cell is one that has the potential to differentiate into any type of human tissue but cannot support the full development of an organism. These cells then become slightly more specialized, and are referred to as multipotent cells.

A **multipotent** stem cell has the potential to differentiate into different types of cells within a given cell lineage or small number of lineages, such as a red blood cell or white blood cell.

Finally, multipotent cells can become further specialized oligopotent cells. An **oligopotent** stem cell is limited to becoming one of a few different cell types. In contrast, a **unipotent** cell is fully specialized and can only reproduce to generate more of its own specific cell type.

Stem cells are unique in that they can also continually divide and regenerate new stem cells instead of further specializing. There are different stem cells present at different stages of a human’s life. They include the embryonic stem cells of the embryo, fetal stem cells of the fetus, and adult stem cells in the adult. One type of adult stem cell is the epithelial stem cell, which gives rise to the keratinocytes in the multiple layers of epithelial cells in the epidermis of skin. Adult bone marrow has three distinct types of stem cells: hematopoietic stem cells, which give rise to red blood cells, white blood cells, and platelets ([link]); endothelial stem cells, which give rise to the endothelial cell types that line blood and lymph vessels; and mesenchymal stem cells, which give rise to the different types of muscle cells.

Hematopoiesis



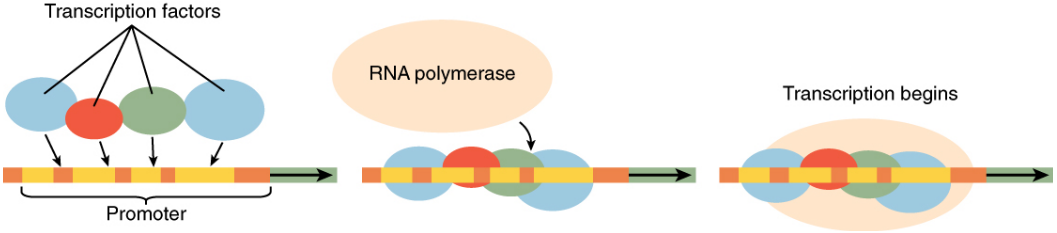
The process of hematopoiesis involves the differentiation of multipotent cells into blood and immune cells. The multipotent hematopoietic stem cells give rise to many different cell types, including the cells of the immune system and red blood cells.

Differentiation

When a cell differentiates (becomes more specialized), it may undertake major changes in its size, shape, metabolic activity, and overall function. Because all cells in the body, beginning with the fertilized egg, contain the same DNA, how do the different cell types come to be so different? The answer is analogous to a movie script. The different actors in a movie all read from the same script, however, they are each only reading their own part of the script. Similarly, all cells contain the same full complement of DNA, but each type of cell only “reads” the portions of DNA that are relevant to its own function. In biology, this is referred to as the unique genetic expression of each cell.

In order for a cell to differentiate into its specialized form and function, it need only manipulate those genes (and thus those proteins) that will be expressed, and not those that will remain silent. The primary mechanism by which genes are turned “on” or “off” is through transcription factors. A **transcription factor** is one of a class of proteins that bind to specific genes on the DNA molecule and either promote or inhibit their transcription ([link]).

Transcription Factors Regulate Gene Expression



While each body cell contains the organism’s entire genome, different cells regulate gene expression with the use of various transcription factors. Transcription factors are proteins that affect the binding of RNA polymerase to a particular gene on the DNA molecule.

Note:

Everyday Connection

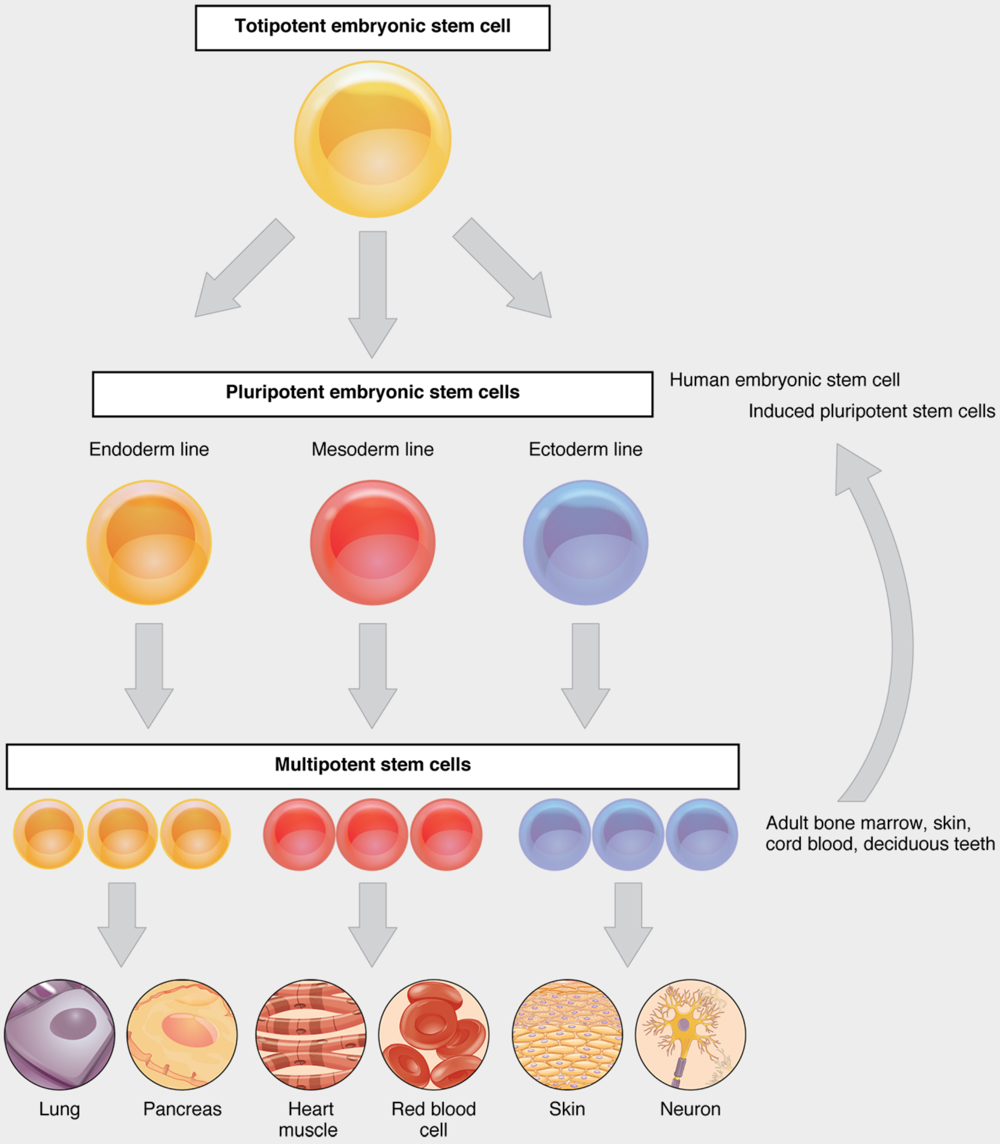
Stem Cell Research

Stem cell research aims to find ways to use stem cells to regenerate and repair cellular damage. Over time, most adult cells undergo the wear and tear of aging and lose their ability to divide and repair themselves. Stem cells do not display a particular morphology or function. Adult stem cells, which exist as a small subset of cells in most tissues, keep dividing and can differentiate into a number of specialized cells generally formed by that tissue. These cells enable the body to renew and repair body tissues.

The mechanisms that induce a non-differentiated cell to become a specialized cell are poorly understood. In a laboratory setting, it is possible to induce stem cells to differentiate into specialized cells by changing the physical and chemical conditions of growth. Several sources of stem cells are used experimentally and are classified according to their origin and potential for differentiation. Human embryonic stem cells (hESCs) are extracted from embryos and are pluripotent. The adult stem cells that are present in many organs and differentiated tissues, such as bone marrow and skin, are multipotent, being limited in differentiation to the types of cells found in those tissues. The stem cells isolated from umbilical cord blood are also multipotent, as are cells from deciduous teeth (baby teeth). Researchers have recently developed induced pluripotent stem cells (iPSCs) from mouse and human adult stem cells. These cells are genetically reprogrammed multipotent adult cells that function like embryonic stem cells; they are capable of generating cells characteristic of all three germ layers.

Because of their capacity to divide and differentiate into specialized cells, stem cells offer a potential treatment for diseases such as diabetes and heart disease ([link]). Cell-based therapy refers to treatment in which stem cells induced to differentiate in a growth dish are injected into a patient to repair damaged or destroyed cells or tissues. Many obstacles must be overcome for the application of cell-based therapy. Although embryonic stem cells have a nearly unlimited range of differentiation potential, they are seen as foreign by the patient’s immune system and may trigger rejection. Also, the destruction of embryos to isolate embryonic stem cells raises considerable ethical and legal questions.

Stem Cells


Totipotent embryonic stem cell
Pluripotent embryonic stem cells
Human embryonic stem cell
Induced pluripotent stem cells
Endoderm line
Mesoderm line
Ectoderm line
Multipotent stem cells
Adult bone marrow, skin, cord blood, deciduous teeth
Lung
Pancreas
Heart muscle
Red blood cell
Skin
Neuron

The capacity of stem cells to differentiate into specialized cells make them potentially valuable in therapeutic applications designed to replace damaged cells of different body tissues.

In contrast, adult stem cells isolated from a patient are not seen as foreign by the body, but they have a limited range of differentiation. Some individuals bank the cord blood or deciduous teeth of their child, storing away those sources of stem cells for future use, should their child need it. Induced pluripotent stem cells are considered a promising advance in the field because using them avoids the legal, ethical, and immunological pitfalls of embryonic stem cells.

Chapter Review

One of the major areas of research in biology is that of how cells specialize to assume their unique structures and functions, since all cells essentially originate from a single fertilized egg. Cell differentiation is the process of cells becoming specialized as they body develops. A stem cell is an unspecialized cell that can divide without limit as needed and can, under specific conditions, differentiate into specialized cells. Stem cells are divided into several categories according to their potential to differentiate. While all somatic cells contain the exact same genome, different cell types only express some of those genes at any given time. These differences in gene expression ultimately dictate a cell’s unique morphological and physiological characteristics. The primary mechanism that determines which genes will be expressed and which ones will not is through the use of different transcription factor proteins, which bind to DNA and promote or hinder the transcription of different genes. Through the action of these transcription factors, cells specialize into one of hundreds of different cell types in the human body.

Review Questions

Exercise:

Problem:

Arrange the following terms in order of increasing specialization: oligopotency, pleuripotency, unipotency, multipotency.

a. multipotency, pleuripotency, oligopotency, unipotency
b. pleuripotency, oligopotency, multipotency unipotency
c. oligopotency, pleuripotency, unipotency, multipotency
d. pleuripotency, multipotency, oligopotency, unipotency

Solution:

D

Exercise:

Problem:

Which type of stem cell gives rise to red and white blood cells?

a. endothelial
b. epithelial
c. hematopoietic
d. mesenchymal

Solution:

C

Exercise:

Problem:

What multipotent stem cells from children sometimes banked by parents?

a. fetal stem cells
b. embryonic stem cells
c. cells from the umbilical cord and from baby teeth
d. hematopoietic stem cells from red and white blood cells

Solution:

C

Critical Thinking Questions

Exercise:

Problem:

Explain how a transcription factor ultimately determines whether or not a protein will be present in a given cell?

Solution:

Transcription factors bind to DNA and either promote or inhibit the transcription of a gene. If they promote the transcription of a particular gene, then that gene will be transcribed and the mRNA subsequently translated into protein. If gene transcription is inhibited, then there will be no way of synthesizing the gene’s corresponding protein.

Exercise:

Problem:

Discuss two reasons why the therapeutic use of embryonic stem cells can present a problem.

Solution:

Embryonic stem cells derive from human embryos, which are destroyed to obtain the cells. The destruction of human embryos is an ethical problem. And, the DNA in an embryonic stem cell would differ from the DNA of the person being treated, which could result in immune problems or rejected of tissue.

Glossary

multipotent	describes the condition of being able to differentiate into different types of cells within a given cell lineage or small number of lineages, such as a red blood cell or white blood cell
oligopotent	

describes the condition of being more specialized than multipotency; the condition of being able to differentiate into one of a few possible cell types

pluripotent
describes the condition of being able to differentiate into a large variety of cell types

stem cell
cell that is oligo-, multi-, or pleuripotent that has the ability to produce additional stem cells rather than becoming further specialized

totipotent
embryonic cells that have the ability to differentiate into any type of cell and organ in the body

transcription factor
one of the proteins that regulate the transcription of genes

unipotent
describes the condition of being committed to a single specialized cell type

Determining the Structure of DNA

By the end of this section, you will be able to:

- Describe the structure of DNA
- Explain the Sanger method of DNA sequencing
- Discuss the similarities and differences between eukaryotic and prokaryotic DNA

The three letters “DNA” have now become associated with crime solving, paternity testing, human identification, and genetic testing. DNA can be retrieved from hair, blood, or saliva. With the exception of identical twins, each person’s DNA is unique and it is possible to detect differences between human beings on the basis of their unique DNA sequence.

DNA analysis has many practical applications beyond forensics and paternity testing. DNA testing is used for tracing genealogy and identifying pathogens. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is now possible to determine predisposition to many diseases by analyzing genes.

DNA is the genetic material passed from parent to offspring for all life on Earth. The technology of molecular genetics developed in the last half century has enabled us to see deep into the history of life to deduce the relationships between living things in ways never thought possible. It also allows us to understand the workings of evolution in populations of organisms. Over a thousand species have had their entire genome sequenced, and there have been thousands of individual human genome sequences completed. These sequences will allow us to understand human disease and the relationship of humans to the rest of the tree of life. Finally, molecular genetics techniques have revolutionized plant and animal breeding for human agricultural needs. All of these advances in biotechnology depended on basic research leading to the discovery of the structure of DNA in 1953, and the research since then that has uncovered the details of DNA replication and the complex process leading to the expression of DNA in the form of proteins in the cell.

Identifying DNA as the Hereditary Agent

In the early Twentieth Century chromosomes were accepted as the structure that carried the hereditary information. However, chemically, chromosomes contained both proteins and DNA. At the time it was widely accepted that it had to be the proteins that carried the hereditary information, because the DNA molecule was considered too simple to account for the the complexity and diversity of cells.

In the late 1920s Frederick Griffith ran a series of experiments with bacteria showing that traits can be passed from one cell to another. In his experiment he used two strains of bacteria one virulent, the other non-virulent. He used these strains to infect mice and recorded the results. As expected when the mice were infected with the virulent strain they died and with the non-virulent strain they lived. However, when he heat killed the virulent strain and mixed the dead bacteria with live bacteria from the non-virulent strain, then injected the mice with living cells cultured from this mixture, the mice died. This demonstrated that something in the dead bacteria was still able to transform non-virulent strains of bacteria to virulent strains. These experiments did not determine what the transforming agent was, but they laid the groundwork for future studies.

In the mid 1940s Oswald Avery, Colin MacLeod, and Maclyn McCarty refined Fredrick Griffith's experiment. Using a series of chemical techniques they designed experiments to test the ability of different molecules in the cells of the dead bacteria to transform non-virulent live bacteria. They found that proteases did not affect the transformation process, but an enzyme that breaks down DNA stopped the transformation. Even though these experiments were replicated successfully in other labs, a large number of scientists continued to believe that DNA was too simple a molecule to carry the information and that proteins, not DNA, carried the hereditary information.

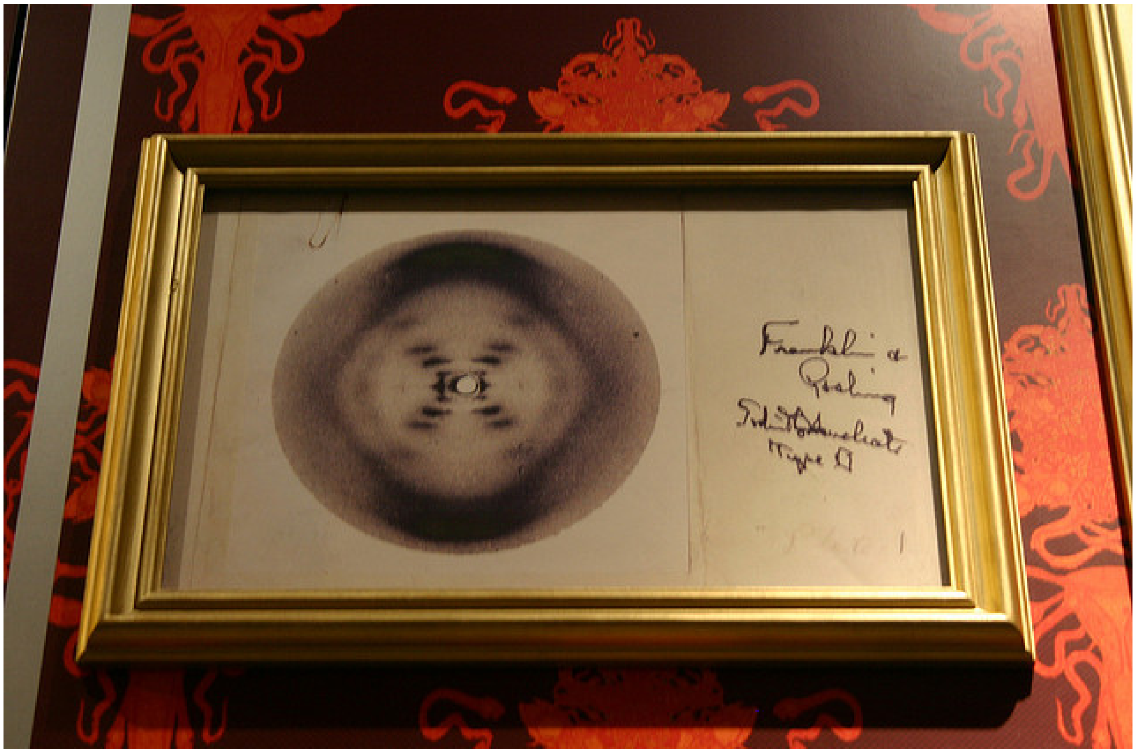
In the early 1950s Alfred Hershey and Martha Chase developed a different experimental design to support the case for DNA as the transforming agent. Bacteriophages are a class of viruses that infect bacteria. These viruses are very simple, consisting of a protein coat and DNA in the core. In order to reproduce they must infect a bacterial cell and transform it into a virus factory. They reasoned that whatever the virus injects into the cell must be the transforming agent. Sulfur is an element found in amino acids but not nucleic acids, and phosphorus is found in nucleic acids but not in amino acids.

Hershey and Chase cultured two groups of viruses, one in a solution containing radioactive isotopes of sulfur, the other in a solution containing radioactive phosphorus. They then infected one bacterial colony with the virus containing radioactive sulfur and another with the radioactive phosphorus. After giving the virus time to infect the cells, they used a high speed blender to remove the bacteriophages from the cell surface. They found that the infected bacteria contained the radioactive phosphorus, but not the radioactive sulfur. They concluded that protein was unlikely to be the transforming agent; however, they stopped short of claiming that DNA was the agent, only that it must play a role.

As the evidence mounted for DNA as the basis of heredity, the search intensified to understand both the precise chemical structure of DNA and its mode of action in determining heredity. A number of scientists played a role. An early proponent of Avery's evidence for DNA as the agent, Erwin Chargaff, set out to develop a reliable technique to precisely measure the percentage of each type of nucleotide in specific DNA samples. He intended to use it to compare the DNA composition of different species. He believed that the diversity of living things could be explained by the sequences of nucleotides much like the sequences of amino acids determined the diversity of proteins. As Chargaff began collecting precise information on nucleotide percentages a pattern quickly emerged. DNA consists of four nucleotides commonly abbreviated as A C T G. The pattern Chargaff found was that while percentages of these nucleotides differed among specimens, the percentages for A and T were always the same within a specimen as were the percentages for G and C. This became known as Chargaff's Rule.

In the 1950s, Francis Crick and James Watson worked together to determine the structure of DNA at the University of Cambridge, England. Other scientists like Linus Pauling and Maurice Wilkins were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. In Wilkins’ lab, researcher Rosalind

Franklin was using X-ray diffraction methods to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule on the basis of Franklin's data because Crick, who also had experience with X-ray diffraction, was able to interpret Franklin's images ([link]). There is some debate as to whether Crick and Watson actually had Franklin's permission to view the images.



Rosalind Franklin's Photo 51

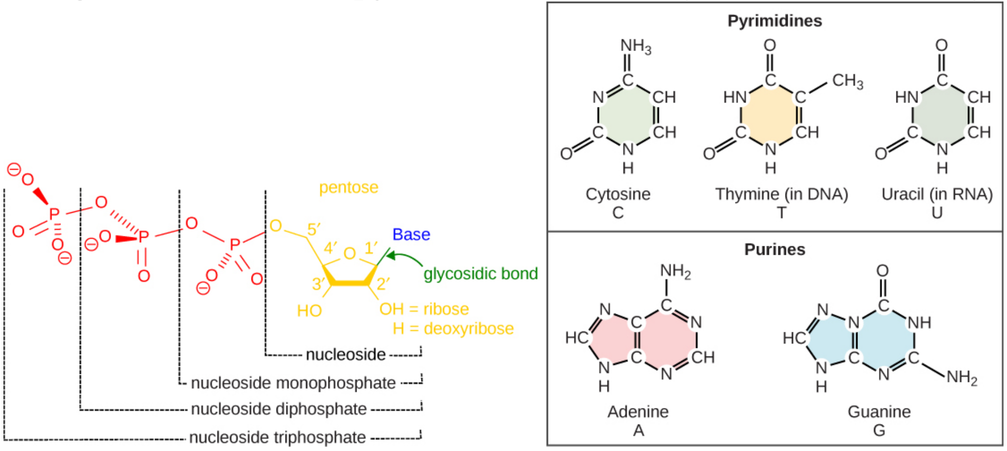
Crick and Watson's structural model of DNA ([link]) was submitted to Nature in 1953 and led to the elucidation of how genetic material is replicated and divided among offspring, launching all of modern molecular biology. It was the first publication which described the discovery of the double helix structure of DNA. This discovery had a major impact on biology, particularly in the field of genetics. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine.

However, much credit is due to Rosalind Franklin's Crystal structure Photo 51 that determined the double helix shape, and for which she never received due credit. Franklin died in 1958 of ovarian cancer at the age of 37 and Nobel prizes are not awarded posthumously.

Crick and Watson with their original DNA model

The Structure of DNA

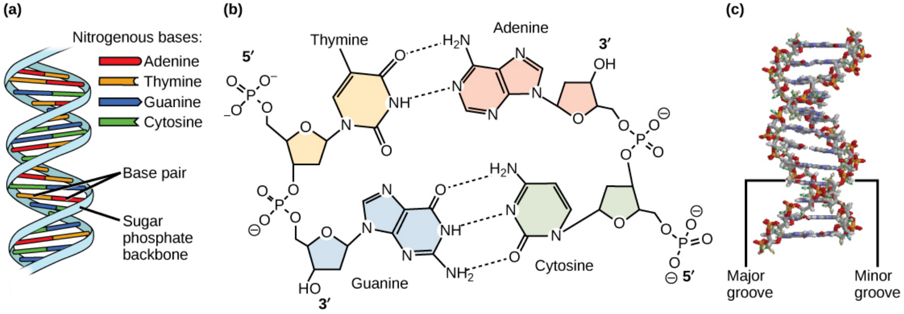
The building blocks of DNA are nucleotides. The important components of the nucleotide are a nitrogenous base, deoxyribose (5-carbon sugar), and a phosphate group ([link]). The nucleotide is named depending on the nitrogenous base. The nitrogenous base can be a purine such as adenine (A) and guanine (G), or a pyrimidine such as cytosine (C) and thymine (T).



Each nucleotide is made up of a sugar, a phosphate group, and a nitrogenous base. The sugar is deoxyribose in DNA and ribose in RNA.

The nucleotides combine with each other by covalent bonds known as phosphodiester bonds or linkages. The purines have a double ring structure with a six-membered ring fused to a five-membered ring. Pyrimidines are smaller in size; they have a single six-membered ring structure. The carbon atoms of the five-carbon sugar are numbered 1', 2', 3', 4', and 5' (1' is read as “one prime”). The phosphate residue is attached to the hydroxyl group of the 5' carbon of one sugar of one nucleotide and the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, thereby forming a 5'-3' phosphodiester bond.

Watson and Crick proposed that DNA is made up of two strands that are twisted around each other to form a right-handed helix. Base pairing takes place between a purine and pyrimidine; namely, A pairs with T and G pairs with C, which was anticipated by Chargaff's results. Adenine and thymine are complementary base pairs, and cytosine and guanine are also complementary base pairs. The base pairs are stabilized by hydrogen bonds; adenine and thymine form two hydrogen bonds and cytosine and guanine form three hydrogen bonds. The two strands are **anti-parallel** in nature; that is, the 3' end of one strand faces the 5' end of the other strand. The sugar and phosphate of the nucleotides form the backbone of the structure, whereas the nitrogenous bases are stacked inside. Each base pair is separated from the other base pair by a distance of 0.34 nm, and each turn of the helix measures 3.4 nm. Therefore, ten base pairs are present per turn of the helix. The diameter of the DNA double helix is 2 nm, and it is uniform throughout. Only the pairing between a purine and pyrimidine can explain the uniform diameter. The twisting of the two strands around each other results in the formation of uniformly spaced major and minor grooves ([link]).



DNA has (a) a double helix structure and (b) phosphodiester bonds. The (c) major and minor grooves are binding sites for DNA binding proteins during processes such as transcription (the copying of RNA from DNA) and replication.

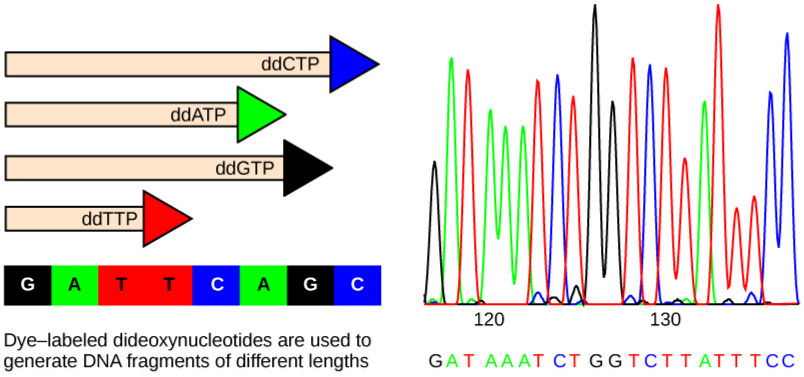
DNA Sequencing Techniques

Until the 1990s, the sequencing of DNA (reading the sequence of DNA) was a relatively expensive and long process. Using radiolabeled nucleotides also compounded the problem through safety concerns. With currently available technology and automated machines, the process is cheap, safer, and can be completed in a matter of hours. Fred Sanger developed the sequencing method used for the human genome sequencing project, which is widely used today ([link]).

Note:
Link to Learning

Visit this site to watch a video explaining the DNA sequence reading technique that resulted from Sanger’s work.

The method is known as the dideoxy chain termination method. The sequencing method is based on the use of chain terminators, the dideoxynucleotides (ddNTPs). The dideoxynucleotides, or ddNTPSs, differ from the deoxynucleotides by the lack of a free 3' OH group on the five-carbon sugar. If a ddNTP is added to a growing a DNA strand, the chain is not extended any further because the free 3' OH group needed to add another nucleotide is not available. By using a predetermined ratio of deoxyribonucleotides to dideoxynucleotides, it is possible to generate DNA fragments of different sizes.



Dye–labeled dideoxynucleotides are used to generate DNA fragments of different lengths

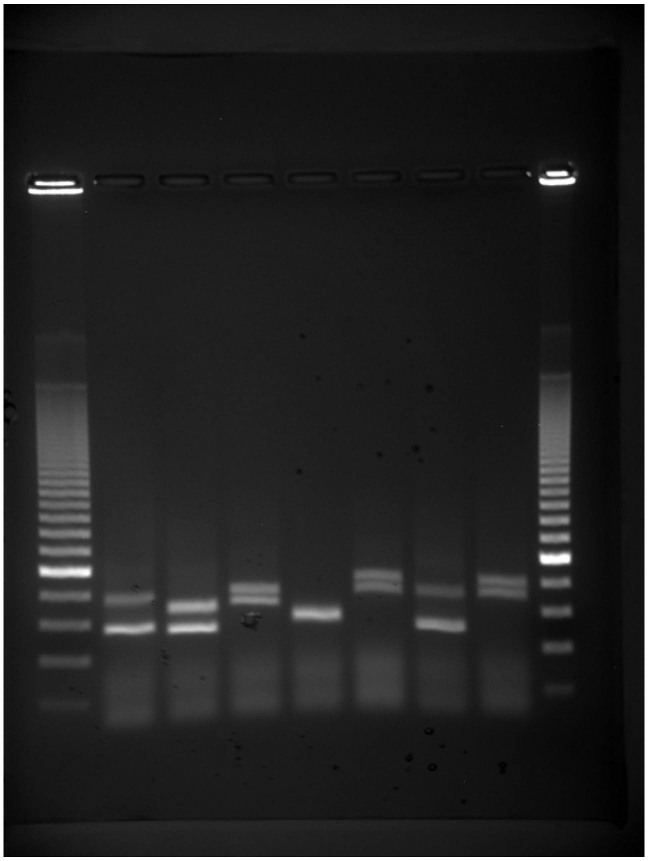
In Frederick Sanger's dideoxy chain termination method, dye-labeled dideoxynucleotides are used to generate DNA fragments that terminate at different points. The DNA is separated by capillary electrophoresis on the basis of size, and from the order of fragments formed, the DNA sequence can be read. The DNA sequence readout is shown on an electropherogram that is generated by a laser scanner.

The DNA sample to be sequenced is denatured or separated into two strands by heating it to high temperatures. The DNA is divided into four tubes in which a primer, DNA polymerase, and all four nucleotides (A, T, G, and C) are added. In addition to each of the four tubes, limited quantities of one of the four dideoxynucleotides are added to each tube respectively. The tubes are labeled as A, T, G, and C according to the ddNTP added. For detection purposes, each of the four dideoxynucleotides carries a different fluorescent label. Chain elongation continues until a fluorescent dideoxy nucleotide is incorporated, after which no further elongation takes place. After the reaction is over, electrophoresis is performed. Even a difference in length of a single base can be detected. The sequence is read from a laser scanner. For his work on DNA sequencing, Sanger received a Nobel Prize in chemistry in 1980.

Note:
Link to Learning

Sanger’s genome sequencing has led to a race to sequence human genomes at a rapid speed and low cost, often referred to as the $1000 in one day sequence. Learn more by selecting the Sequencing at Speed animation here.

Gel **electrophoresis** is a technique used to separate DNA fragments of different sizes. Usually the gel is made of a chemical called agarose. Agarose powder is added to a buffer and heated. After cooling, the gel solution is poured into a casting tray. Once the gel has solidified, the DNA is loaded on the gel and electric current is applied. The DNA has a net negative charge and moves from the negative electrode toward the positive electrode. The electric current is applied for sufficient time to let the DNA separate according to size; the smallest fragments will be farthest from the well (where the DNA was loaded), and the heavier molecular weight fragments will be closest to the well. Once the DNA is separated, the gel is stained with a DNA-specific dye for viewing it ([link]).



DNA can be separated on the basis of size using gel electrophoresis. (credit: James Jacob, Tompkins Cortland Community College)

Note:

Evolution Connection

Neanderthal Genome: How Are We Related?

The first draft sequence of the Neanderthal genome was recently published by Richard E. Green et al. in 2010.[footnote] Neanderthals are the closest ancestors of present-day humans. They were known to have lived in Europe and Western Asia before they disappeared from fossil records approximately 30,000 years ago. Green's team studied almost 40,000-year-old fossil remains that were selected from sites across the world. Extremely sophisticated means of sample preparation and DNA sequencing were employed because of the fragile nature of the bones and heavy microbial contamination. In their study, the scientists were able to sequence some four billion base pairs. The Neanderthal sequence was compared with that of present-day humans from across the world. After comparing the sequences, the researchers found that the Neanderthal genome had 2 to 3 percent greater similarity to people living outside Africa than to people in Africa. While current theories have suggested that all present-day humans can be traced to a small ancestral population in Africa, the data from the Neanderthal genome may contradict this view. Green and his colleagues also discovered DNA segments among people in Europe and Asia that are more similar to Neanderthal sequences than to other contemporary human sequences. Another interesting observation was that Neanderthals are as closely related to people from Papua New Guinea as to those from China or France. This is surprising because Neanderthal fossil remains have been located only in Europe and West Asia. Most likely, genetic exchange took place between Neanderthals and modern humans as modern humans emerged out of Africa, before the divergence of Europeans, East Asians, and Papua New Guineans.

Richard E. Green et al., "A Draft Sequence of the Neandertal Genome," *Science* 328 (2010): 710-22.

Several genes seem to have undergone changes from Neanderthals during the evolution of present-day humans. These genes are involved in cranial structure, metabolism, skin morphology, and cognitive development. One of the genes that is of particular interest is *RUNX2*, which is different in modern day humans and Neanderthals. This gene is responsible for the prominent frontal bone, bell-shaped rib cage, and dental differences seen in Neanderthals. It is speculated that an evolutionary change in *RUNX2* was important in the origin of modern-day humans, and this affected the cranium and the upper body.

Note:

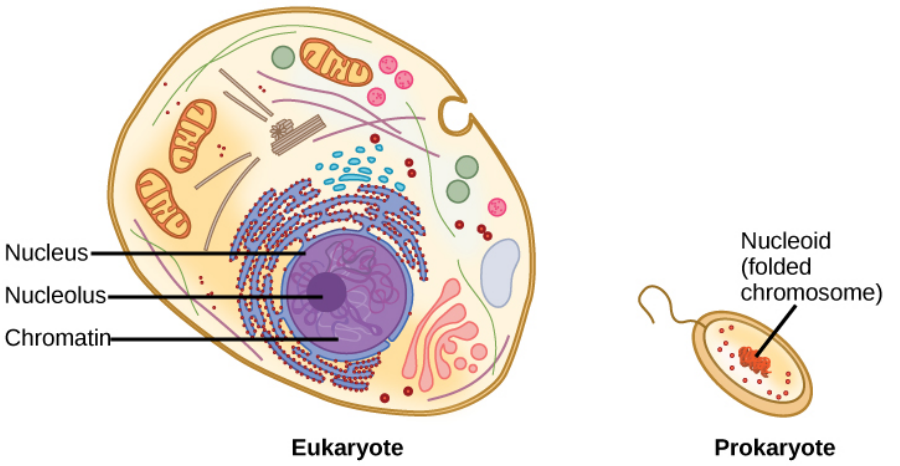
Link to Learning

Watch Svante Pääbo’s talk explaining the Neanderthal genome research at the 2011 annual TED (Technology, Entertainment, Design) conference.

DNA Packaging in Cells

When comparing prokaryotic cells to eukaryotic cells, prokaryotes are much simpler than eukaryotes in many of their features ([link]). Most prokaryotes contain a single, circular chromosome that is found in an area of the cytoplasm called the nucleoid.

Note:
Art Connection



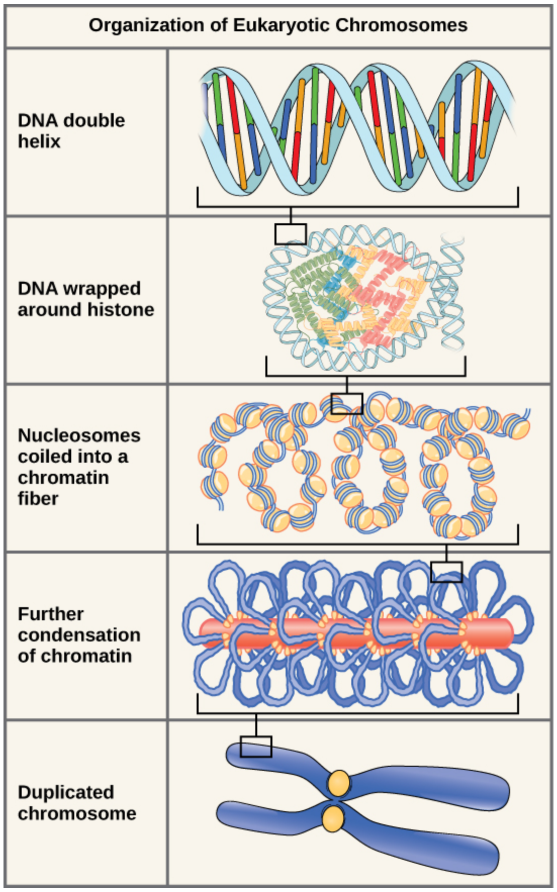
A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

In eukaryotic cells, DNA and RNA synthesis occur in a separate compartment from protein synthesis. In prokaryotic cells, both processes occur together. What advantages might there be to separating the processes? What advantages might there be to having them occur together?

The size of the genome in one of the most well-studied prokaryotes, *E.coli,* is 4.6 million base pairs (approximately 1.1 mm, if cut and stretched out). So how does this fit inside a small bacterial cell? The DNA is twisted by what is known as supercoiling. Supercoiling means that DNA is either under-wound (less than one turn of the helix per 10 base pairs) or over-wound (more than 1 turn per 10 base pairs) from its normal relaxed state. Some proteins are known to be involved in the supercoiling; other proteins and enzymes such as DNA gyrase help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus ([link]). At the most basic level, DNA is wrapped around proteins known as **histones** to form structures called **nucleosomes**. The histones are evolutionarily conserved proteins that are rich in basic amino acids and form an octamer. The DNA (which is negatively charged because of the phosphate groups) is wrapped tightly around the histone core. This nucleosome is linked to the next one with the help of a linker DNA. This is also known as the “beads on a string” structure. This is further compacted into a 30 nm fiber, which is the diameter of the structure. At the metaphase stage, the chromosomes are at their most compact, are approximately 700 nm in width, and are found in association with scaffold proteins.

In interphase, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. The tightly packaged region is known as **heterochromatin**, and the less dense region is known as **euchromatin**. Heterochromatin usually contains genes that are not expressed, and is found in the regions of the centromere and telomeres. The euchromatin usually contains genes that are transcribed, with DNA packaged around nucleosomes but not further compacted.


Organization of Eukaryotic Chromosomes
DNA double helix
DNA wrapped around histone
Nucleosomes coiled into a chromatin fiber
Further condensation of chromatin
Duplicated chromosome

These figures illustrate the compaction of the eukaryotic chromosome.

Section Summary

The currently accepted model of the double-helix structure of DNA was proposed by Watson and Crick. Some of the salient features are that the two strands that make up the double helix are complementary and anti-parallel in nature. Deoxyribose sugars and phosphates form the backbone of the structure, and the nitrogenous bases are stacked inside. The diameter of the double helix, 2 nm, is uniform throughout. A purine always pairs with a pyrimidine; A pairs with T, and G pairs with C. One turn of the helix has ten base pairs.

During cell division, each daughter cell receives a copy of the DNA by a process known as DNA replication. The model for DNA replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied. The semi-conservative method suggests that each of the two parental DNA strands acts as template for new DNA to be synthesized; after replication, each double-stranded DNA includes one parental or “old” strand and one “new” strand.

In general, eukaryotic chromosomes contain a linear DNA molecule packaged into nucleosomes, and have two distinct regions that can be distinguished by staining, reflecting different states of packaging and compaction.

Art Connections

Exercise:

Problem:

[link] In eukaryotic cells, DNA and RNA synthesis occur in a separate compartment from protein synthesis. In prokaryotic cells, both processes occur together. What advantages might there be to separating the processes? What advantages might there be to having them occur together?

Solution:

[link] Compartmentalization enables a eukaryotic cell to divide processes into discrete steps so it can build more complex protein and RNA products. But there is an advantage to having a single compartment as well: RNA and protein synthesis occurs much more quickly in a prokaryotic cell.

Review Questions

Exercise:

Problem:DNA double helix does not have which of the following?

a. antiparallel configuration
b. complementary base pairing
c. major and minor grooves
d. uracil

Solution:

D

Exercise:

Problem:

If the sequence of the 5'-3' strand is AATGCTAC, then the complementary sequence has the following sequence:

a. 3'-AATGCTAC-5'
b. 3'-CATCGTAA-5'
c. 3'-TTACGATG-5'
d. 3'-GTAGCATT-5'

Solution:

C

Exercise:

Problem: In eukaryotes, what is the DNA wrapped around?

a. single-stranded binding proteins
b. sliding clamp
c. polymerase
d. histones

Solution:

D

Free Response

Exercise:

Problem: Provide a brief summary of the Sanger sequencing method.

Solution:

The template DNA strand is mixed with a DNA polymerase, a primer, the 4 deoxynucleotides, and a limiting concentration of 4 dideoxynucleotides. DNA polymerase synthesizes a strand complementary to the template. Incorporation of ddNTPs at different locations results in DNA fragments that have terminated at every possible base in the template. These fragments are separated by gel electrophoresis and visualized by a laser detector to determine the sequence of bases.

Exercise:

Problem:

Describe the structure and complementary base pairing of DNA.

Solution:

DNA has two strands in anti-parallel orientation. The sugar-phosphate linkages form a backbone on the outside, and the bases are paired on the inside: A with T, and G with C, like rungs on a spiral ladder.

Glossary

electrophoresis

technique used to separate DNA fragments according to size

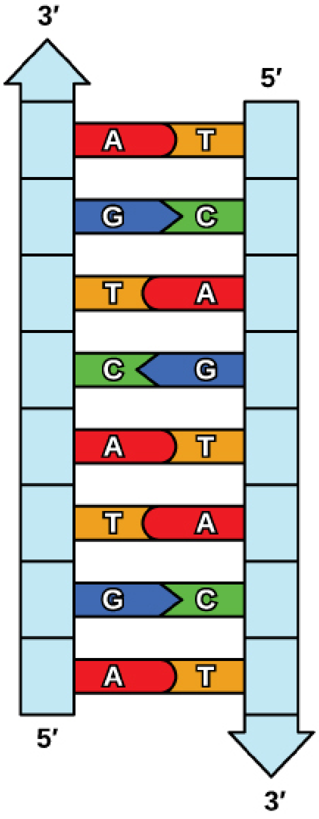
DNA Replication

By the end of this section, you will be able to:

- Explain the process of DNA replication
- Explain the importance of telomerase to DNA replication
- Describe mechanisms of DNA repair

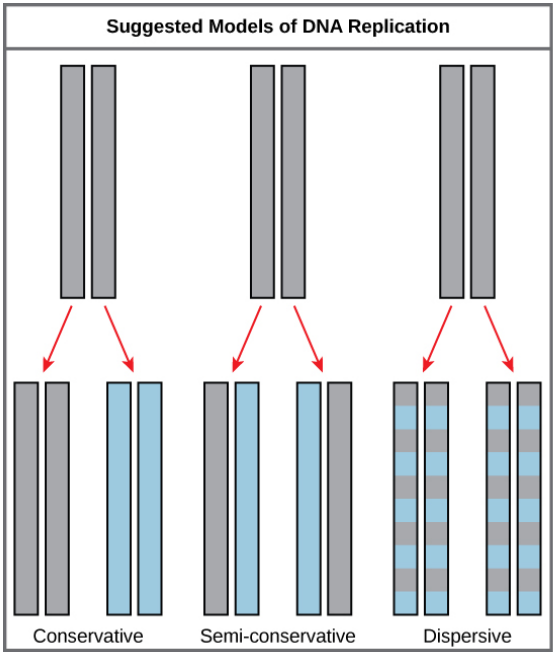
When a cell divides, it is important that each daughter cell receives an identical copy of the DNA. This is accomplished by the process of DNA replication. The replication of DNA occurs during the synthesis phase, or S phase, of the cell cycle, before the cell enters mitosis or meiosis.

The elucidation of the structure of the double helix provided a hint as to how DNA is copied. Recall that adenine nucleotides pair with thymine nucleotides, and cytosine with guanine. This means that the two strands are complementary to each other. For example, a strand of DNA with a nucleotide sequence of AGTCATGA will have a complementary strand with the sequence TCAGTACT ([link]).



The two strands of DNA are complementary, meaning the sequence of bases in one strand can be used to create the correct sequence of bases in the other strand.

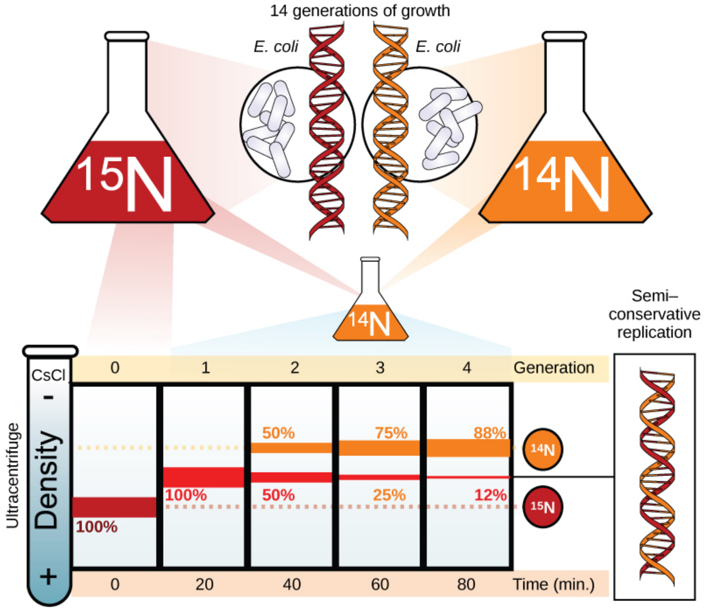
Because of the complementarity of the two strands, having one strand means that it is possible to recreate the other strand. This model for replication suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied. What was not clear was how the replication took place. There were three models suggested ([link]): conservative, semi-conservative, and dispersive.



The three suggested models of DNA replication. Grey indicates the original DNA strands, and blue indicates newly synthesized

In conservative replication, the parental DNA remains together, and the newly formed daughter strands are together. The semi-conservative method suggests that each of the two parental DNA strands act as a template for new DNA to be synthesized; after replication, each double-stranded DNA includes one parental or “old” strand and one “new” strand. In the dispersive model, both copies of DNA have double-stranded segments of parental DNA and newly synthesized DNA interspersed.

Matthew Meselson and Franklin Stahl were interested in understanding how DNA replicates. They grew *E. coli* for several generations in a medium containing a “heavy” isotope of nitrogen (^{15}N) that gets incorporated into nitrogenous bases, and eventually into the DNA ([link]).



Meselson and Stahl experimented with *E. coli* grown first in heavy nitrogen

The *E. coli* culture was then shifted into a medium containing ^{14}N and allowed to grow for one generation. The cells were harvested and the DNA was isolated. The DNA was centrifuged at high speeds in an ultracentrifuge. Some cells were allowed to grow for one more life cycle in ^{14}N and spun again. During the density gradient centrifugation, the DNA is loaded into a gradient (typically a salt such as cesium chloride or sucrose) and spun at high speeds of 50,000 to 60,000 rpm. Under these circumstances, the DNA will form a band according to its density in the gradient. DNA grown in ^{15}N will band at a higher density position than that grown in ^{14}N. Meselson and Stahl noted that after one generation of growth in ^{14}N after they had been shifted from ^{15}N, the single band observed was intermediate in position in between DNA of cells grown exclusively in ^{15}N and ^{14}N. This suggested either a semi-conservative or dispersive mode of replication. The DNA harvested from cells grown for two generations in ^{14}N formed two bands: one DNA band was at the intermediate position between ^{15}N and ^{14}N, and the other corresponded to the band of ^{14}N DNA. These results could only be explained if DNA replicates in a semi-conservative manner. Therefore, the other two modes were ruled out.

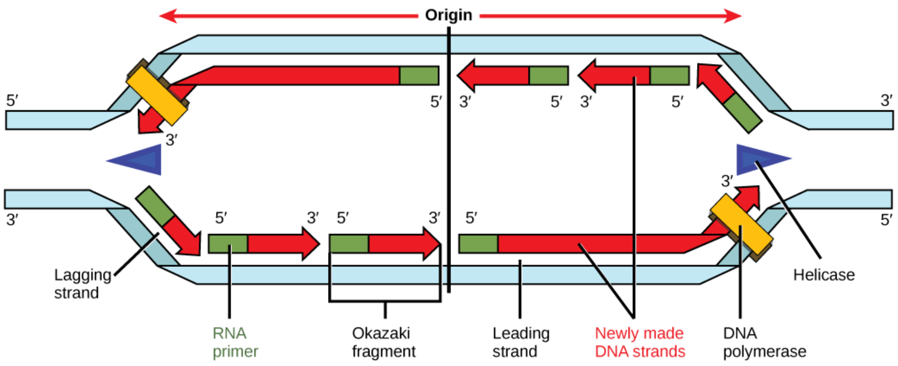
The DNA Replication Process

DNA replication is a very complicated process that involves several enzymes and other proteins. It occurs in three main stages: initiation, elongation, and termination.

Stage 1: Initiation. The two complementary strands are separated, much like unzipping a zipper. Special enzymes, including **helicase**, untwist and separate the two strands of DNA.

Stage 2: Elongation. Each strand becomes a template along which a new complementary strand is built. **DNA polymerase** brings in the correct bases to complement the template strand, synthesizing a new strand base by base. A DNA polymerase is an enzyme that adds free nucleotides to the end of a chain of DNA, making a new double strand. This growing strand continues to be built until it has fully complemented the template strand.

Stage 3: Termination. Once the two original strands are bound to their own, finished, complementary strands, DNA replication is stopped and the two new identical DNA molecules are complete.



A replication fork is formed by the opening of the origin of replication, and helicase separates the DNA strands. An RNA primer is synthesized, and is elongated by the DNA polymerase. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches. The DNA fragments are joined by DNA ligase (not shown).

Recall that eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. How does the replication machinery know where on the DNA double helix to begin? It turns out that there are specific nucleotide sequences called **origins of replication** at which replication begins. Certain proteins bind to the origin of replication while an enzyme called **helicase** unwinds and opens up the DNA helix. As the DNA opens up, Y-shaped structures called **replication forks** are formed ([link]). Two replication forks are formed at the origin of replication, and these get extended in both directions as replication proceeds. There are multiple origins of replication on the eukaryotic chromosome, such that replication can occur simultaneously from several places in the genome.

During elongation, an enzyme called **DNA polymerase** adds DNA nucleotides to the 3' end of the template. Because DNA polymerase can only add new nucleotides at the end of a backbone, a **primer** sequence, which provides this starting point, is added with complementary RNA nucleotides. This primer is removed later, and the nucleotides are replaced with DNA nucleotides. One strand, which is complementary to the parental DNA strand, is synthesized continuously toward the replication fork so the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. Because DNA polymerase can only synthesize DNA in a 5' to 3' direction, the other new strand is put together in short pieces called **Okazaki fragments**. The Okazaki fragments each require a primer made of RNA to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**. As synthesis proceeds, an enzyme removes the RNA primer, which is then replaced with DNA nucleotides, and the gaps between fragments are sealed by an enzyme called **DNA ligase**.

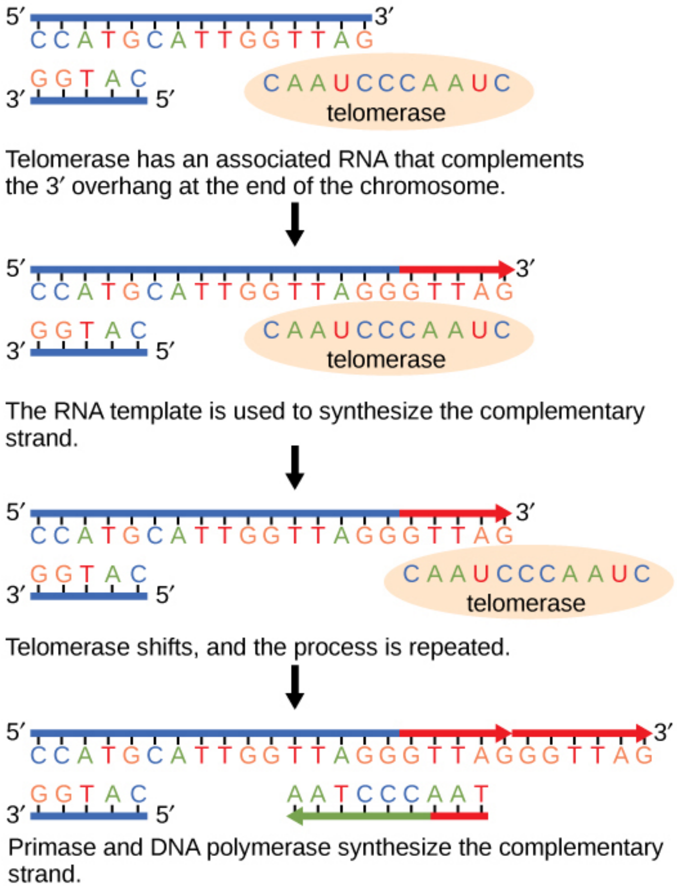
This process continues until the cell’s entire **genome**, the entire complement of an organism’s DNA, is replicated. As you might imagine, it is very important that DNA replication take place precisely so that new cells in the body contain the exact same genetic material as their parent cells.

Note:
Link to Learning

Click through this tutorial on DNA replication.

Telomere Replication

Because eukaryotic chromosomes are linear, DNA replication comes to the end of a line in eukaryotic chromosomes. As you have learned, the DNA polymerase enzyme can add nucleotides in only one direction. In the leading strand, synthesis continues until the end of the chromosome is reached; however, on the lagging strand there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. This presents a problem for the cell because the ends remain unpaired, and over time these ends get progressively shorter as cells continue to divide. The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that do not code for a particular gene. As a consequence, it is telomeres that are shortened with each round of DNA replication instead of genes. For example, in humans, a six base-pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** ([link]) helped in the understanding of how chromosome ends are maintained. The telomerase attaches to the end of the chromosome, and complementary bases to the RNA template are added on the end of the DNA strand. Once the lagging strand template is sufficiently elongated, DNA polymerase can now add nucleotides that are complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.


5′
3′
CCATGCATTGGTTAG
GGTAC
3′
5′
CAAUCCCAAUC
telomerase
Telomerase has an associated RNA that complements the 3′ overhang at the end of the chromosome.
CCATGCATTGGTTAGGGTTAG
The RNA template is used to synthesize the complementary strand.
Telomerase shifts, and the process is repeated.
CCATGCATTGGTTAGGGTTAGGGTTAG
AATCCCAAT
Primase and DNA polymerase synthesize the complementary strand.

The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Telomerase is typically found to be active in germ cells, adult stem cells, and some cancer cells. For her discovery of telomerase and its action, Elizabeth Blackburn ([link]) received the Nobel Prize for Medicine and Physiology in 2009.

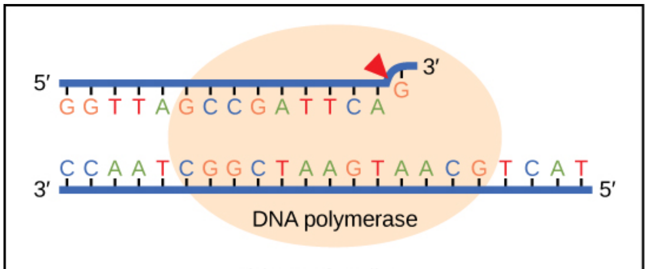
Elizabeth Blackburn, 2009 Nobel Laureate, was the scientist who discovered how telomerase works. (credit: U.S. Embassy, Stockholm, Sweden)

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This essentially means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and this may have potential in regenerative medicine. [footnote] Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

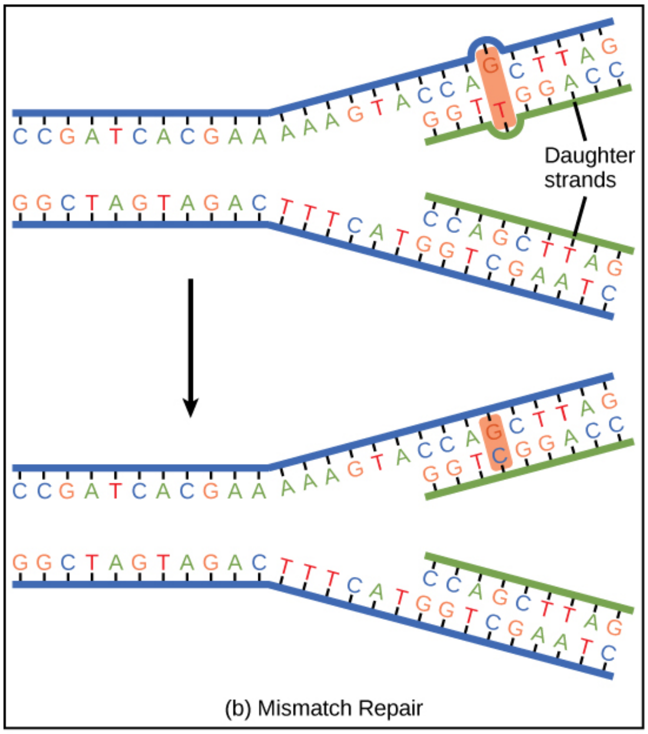
Mariella Jaskelioff, et al., “Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice,” *Nature*, 469 (2011):102–7.

DNA Repair

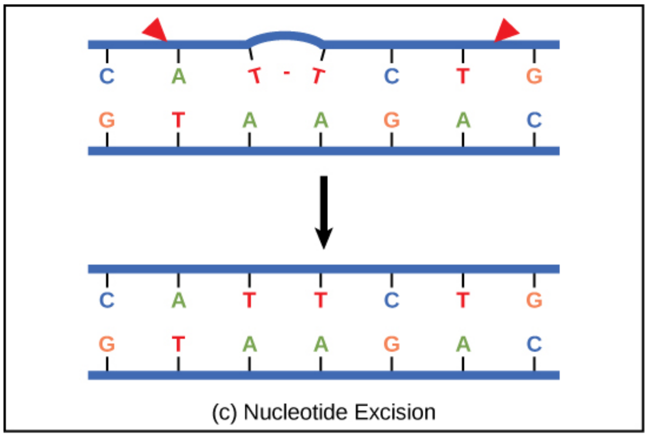
DNA polymerase can make mistakes while adding nucleotides. It edits the DNA by proofreading every newly added base. Incorrect bases are removed and replaced by the correct base, and then polymerization continues ([link]**a**). Most mistakes are corrected during replication, although when this does not happen, the mismatch repair mechanism is employed. Mismatch repair enzymes recognize the wrongly incorporated base and excise it from the DNA, replacing it with the correct base ([link]**b**). In yet another type of repair, nucleotide excision repair, the DNA double strand is unwound and separated, the incorrect bases are removed along with a few bases on the 5' and 3' end, and these are replaced by copying the template with the help of DNA polymerase ([link]**c**). Nucleotide excision repair is particularly important in correcting thymine dimers, which are primarily caused by ultraviolet light. In a thymine dimer, two thymine nucleotides adjacent to each other on one strand are covalently bonded to each other rather than their complementary bases. If the dimer is not removed and repaired it will lead to a mutation. Individuals with flaws in their nucleotide excision repair genes show extreme sensitivity to sunlight and develop skin cancers early in life.



(a) Proofreading



(b) Mismatch Repair



(c) Nucleotide Excision

Proofreading by DNA polymerase (a) corrects errors during replication. In mismatch repair (b), the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base. Nucleotide excision (c) repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

Most mistakes are corrected; if they are not, they may result in a **mutation**—defined as a permanent change in the DNA sequence. Mutations in repair genes may lead to serious consequences like cancer. Mutations in other gene coding regions can result in misshapen proteins unable to perform their normal function. This results in hereditary disorders like Cystic Fibrosis.

Protein Synthesis

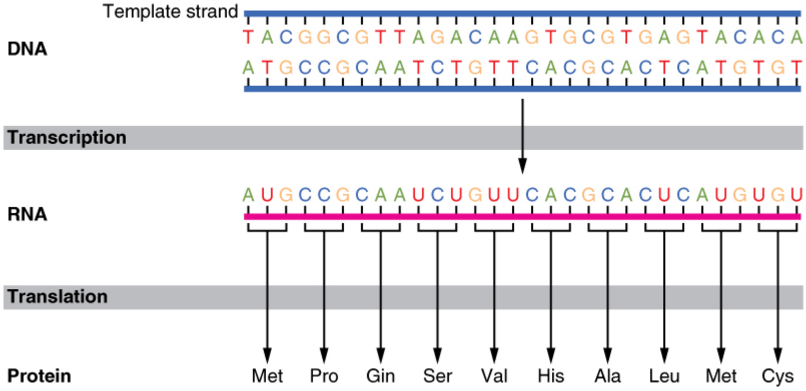
By the end of this section, you will be able to:

- Explain how the genetic code stored within DNA determines the protein that will form
- Describe the process of transcription
- Describe the process of translation
- Discuss the function of ribosomes

It was mentioned earlier that DNA provides a “blueprint” for the cell structure and physiology. This refers to the fact that DNA contains the information necessary for the cell to build one very important type of molecule: the protein. Most structural components of the cell are made up, at least in part, by proteins and virtually all the functions that a cell carries out are completed with the help of proteins. One of the most important classes of proteins is enzymes, which help speed up necessary biochemical reactions that take place inside the cell. Some of these critical biochemical reactions include building larger molecules from smaller components (such as occurs during DNA replication or synthesis of microtubules) and breaking down larger molecules into smaller components (such as when harvesting chemical energy from nutrient molecules). Whatever the cellular process may be, it is almost sure to involve proteins. Just as the cell’s genome describes its full complement of DNA, a cell’s **proteome** is its full complement of proteins. Protein synthesis begins with genes. A **gene** is a functional segment of DNA that provides the genetic information necessary to build a protein. Each particular gene provides the code necessary to construct a particular protein. **Gene expression**, which transforms the information coded in a gene to a final gene product, ultimately dictates the structure and function of a cell by determining which proteins are made.

The interpretation of genes works in the following way. Recall that proteins are polymers, or chains, of many amino acid building blocks. The sequence of bases in a gene (that is, its sequence of A, T, C, G nucleotides) translates to an amino acid sequence. A **triplet** is a section of three DNA bases in a row that codes for a specific amino acid. Similar to the way in which the three-letter code *d-o-g* signals the image of a dog, the three-letter DNA base code signals the use of a particular amino acid. For example, the DNA triplet CAC (cytosine, adenine, and cytosine) specifies the amino acid valine. Therefore, a gene, which is composed of multiple triplets in a unique sequence, provides the code to build an entire protein, with multiple amino acids in the proper sequence ([link]). The mechanism by which cells turn the DNA code into a protein product is a two-step process, with an RNA molecule as the intermediate.

The Transcription/Translation Process


Template strand
DNA
TACGGCGTTAGACAAGTGCGTGAGTACACA
ATGCCGCAATCTGTTCACGCACTCATGTGT
Transcription
RNA
AUGCCGCAAUCUGUUCACGCACUCAUGUGU
Translation
Protein
Met
Pro
Gin
Ser
Val
His
Ala
Leu
Met
Cys

DNA holds all of the genetic information necessary to build a cell’s proteins. The nucleotide sequence of a gene is ultimately translated into an amino acid sequence of the gene’s corresponding protein.

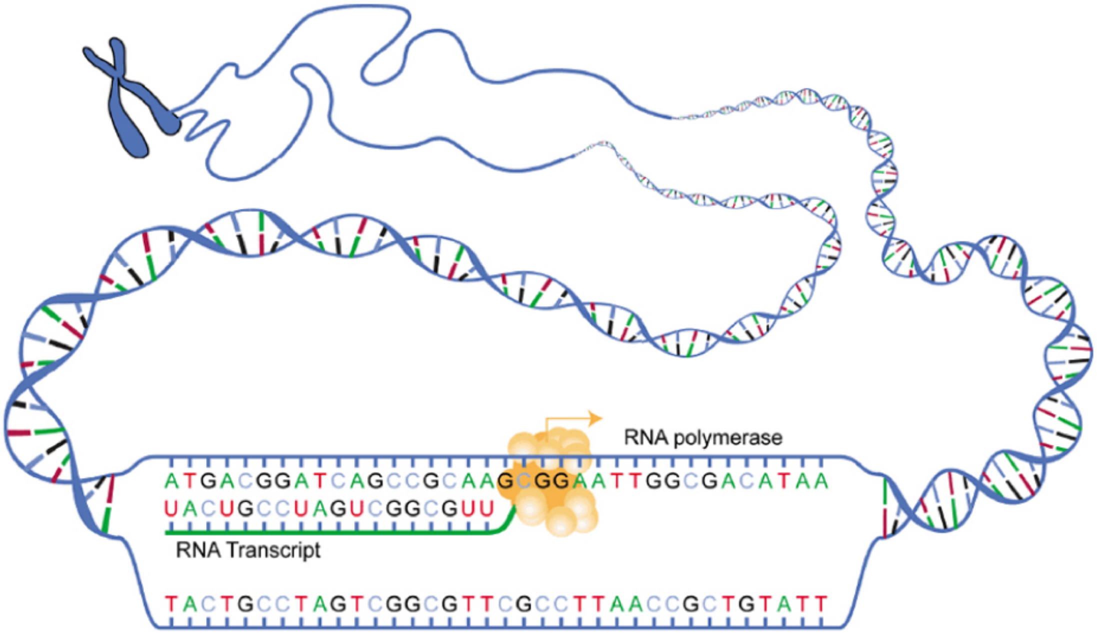
From DNA to RNA: Transcription

DNA is housed within the nucleus, and protein synthesis takes place in the cytoplasm, thus there must be some sort of intermediate messenger that leaves the nucleus and manages protein synthesis. This intermediate messenger is **messenger RNA (mRNA)**, a single-stranded nucleic acid that carries a copy of the genetic code for a single gene out of the nucleus and into the cytoplasm where it is used to produce proteins.

There are several different types of RNA, each having different functions in the cell. The structure of RNA is similar to DNA with a few small exceptions. For one thing, unlike DNA, most types of RNA, including mRNA, are single-stranded and contain no complementary strand. Second, the ribose sugar in RNA contains an additional oxygen atom compared with DNA. Finally, instead of the base thymine, RNA contains the base uracil. This means that adenine will always pair up with uracil during the protein synthesis process.

Gene expression begins with the process called **transcription**, which is the synthesis of a strand of mRNA that is complementary to the gene of interest. This process is called transcription because the mRNA is like a transcript, or copy, of the gene’s DNA code. Transcription begins in a fashion somewhat like DNA replication, in that a region of DNA unwinds and the two strands separate, however, only that small portion of the DNA will be split apart. The triplets within the gene on this section of the DNA molecule are used as the template to transcribe the complementary strand of RNA ([link]). A **codon** is a three-base sequence of mRNA, so-called because they directly encode amino acids. Like DNA replication, there are three stages to transcription: initiation, elongation, and termination.

Transcription: from DNA to mRNA


RNA polymerase
ATGACGGATCAGCCGCAAGCGGAATTGGCGACATAA
UACUGCCUAGUCGGCGUU
RNA Transcript
TACTGCCTAGTCGGCGTTCGCCTTAACCGCTGTATT

In the first of the two stages of making protein from DNA, a gene on the DNA molecule is transcribed into a complementary mRNA molecule.

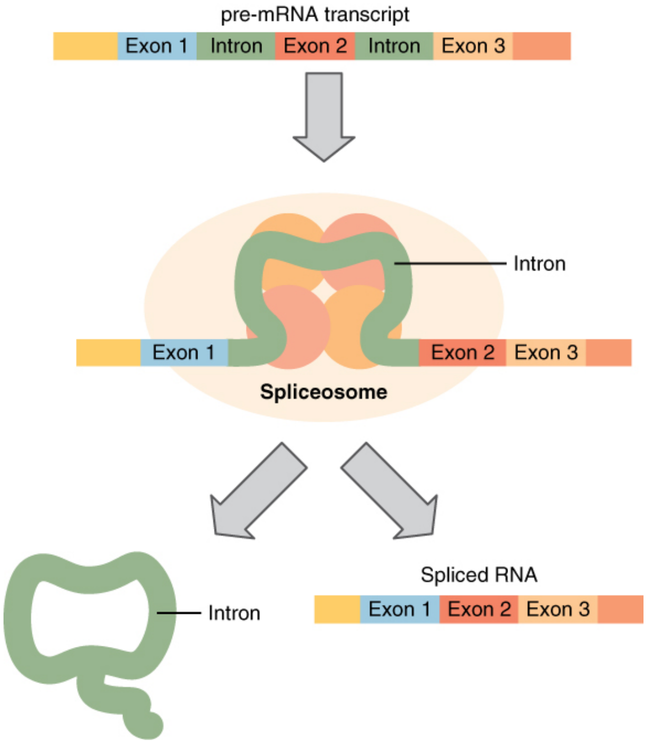
Stage 1: Initiation. A region at the beginning of the gene called a **promoter**—a particular sequence of nucleotides—triggers the start of transcription.

Stage 2: Elongation. Transcription starts when RNA polymerase unwinds the DNA segment. One strand, referred to as the coding strand, becomes the template with the genes to be coded. The polymerase then aligns the correct nucleic acid (A, C, G, or U) with its complementary base on the coding strand of DNA. **RNA polymerase** is an enzyme that adds new nucleotides to a growing strand of RNA. This process builds a strand of mRNA.

Stage 3: Termination. When the polymerase has reached the end of the gene, one of three specific triplets (UAA, UAG, or UGA) codes a "stop" signal, which triggers the enzymes to terminate transcription and release the mRNA transcript.

Before the mRNA molecule leaves the nucleus and proceeds to protein synthesis, it is modified in a number of ways. For this reason, it is often called a pre-mRNA at this stage. For example, your DNA, and thus complementary mRNA, contains long regions called non-coding regions that do not code for amino acids. Their function is still a mystery, but the process called **splicing** removes these non-coding regions from the pre-mRNA transcript ([link]). A **spliceosome**—a structure composed of various proteins and other molecules—attaches to the mRNA and "splices" or cuts out the non-coding regions. The removed segment of the transcript is called an **intron**. The remaining exons are pasted together. An **exon** is a segment of RNA that remains after splicing. Interestingly, some introns that are removed from mRNA are not always non-coding. When different coding regions of mRNA are spliced out, different variations of the protein will eventually result, with differences in structure and function. This process results in a much larger variety of possible proteins and protein functions. When the mRNA transcript is ready, it travels out of the nucleus and into the cytoplasm.

Splicing DNA



In the nucleus, a structure called a spliceosome cuts out introns (noncoding regions) within a pre-mRNA transcript and reconnects the exons.

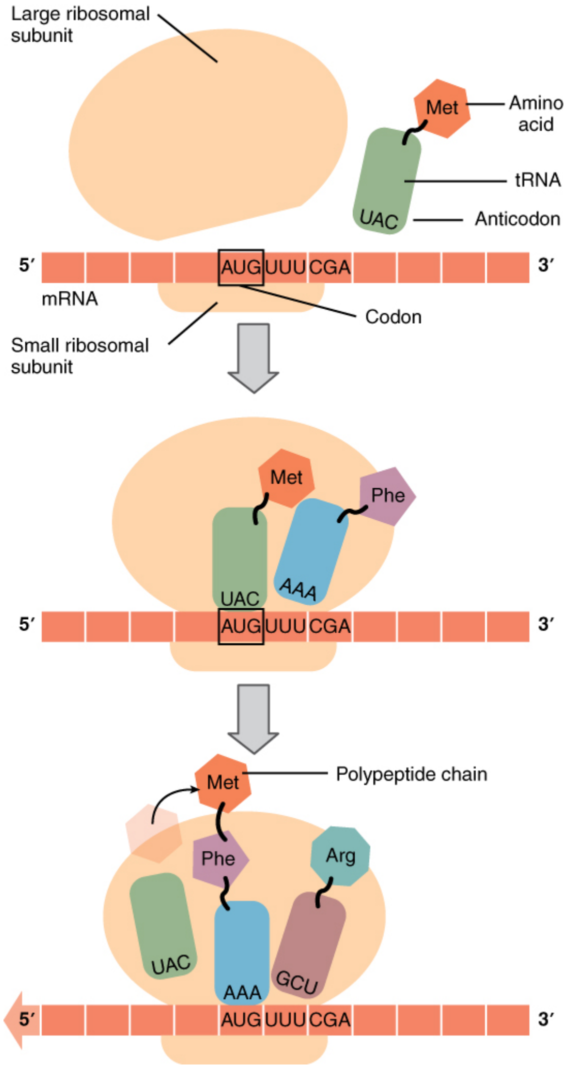
From RNA to Protein: Translation

Like translating a book from one language into another, the codons on a strand of mRNA must be translated into the amino acid alphabet of proteins. **Translation** is the process of synthesizing a chain of amino acids called a **polypeptide**. Translation requires two major aids: first, a "translator," the molecule that will conduct the translation, and second, a substrate on which the mRNA strand is translated into a new protein, like the translator's "desk." Both of these requirements are fulfilled by other types of RNA. The substrate on which translation takes place is the ribosome.

Remember that many of a cell’s ribosomes are found associated with the rough ER, and carry out the synthesis of proteins destined for the Golgi apparatus. **Ribosomal RNA (rRNA)** is a type of RNA that, together with proteins, composes the structure of the ribosome. Ribosomes exist in the cytoplasm as two distinct components, a small and a large subunit. When an mRNA molecule is ready to be translated, the two subunits come together and attach to the mRNA. The ribosome provides a substrate for translation, bringing together and aligning the mRNA molecule with the molecular “translators” that must decipher its code.

The other major requirement for protein synthesis is the translator molecules that physically “read” the mRNA codons. **Transfer RNA (tRNA)** is a type of RNA that ferries the appropriate corresponding amino acids to the ribosome, and attaches each new amino acid to the last, building the polypeptide chain one-by-one. Thus tRNA transfers specific amino acids from the cytoplasm to a growing polypeptide. The tRNA molecules must be able to recognize the codons on mRNA and match them with the correct amino acid. The tRNA is modified for this function. On one end of its structure is a binding site for a specific amino acid. On the other end is a base sequence that matches the codon specifying its particular amino acid. This sequence of three bases on the tRNA molecule is called an **anticodon**. For example, a tRNA responsible for shuttling the amino acid glycine contains a binding site for glycine on one end. On the other end it contains an anticodon that complements the glycine codon (GGA is a codon for glycine, and so the tRNAs anticodon would read CCU). Equipped with its particular cargo and matching anticodon, a tRNA molecule can read its recognized mRNA codon and bring the corresponding amino acid to the growing chain ([link]).

Translation from RNA to Protein



During translation, the mRNA transcript is “read” by a functional complex consisting of the ribosome and tRNA molecules. tRNAs bring the appropriate amino acids in sequence to the growing polypeptide chain by matching their anti-codons with codons on the mRNA strand.

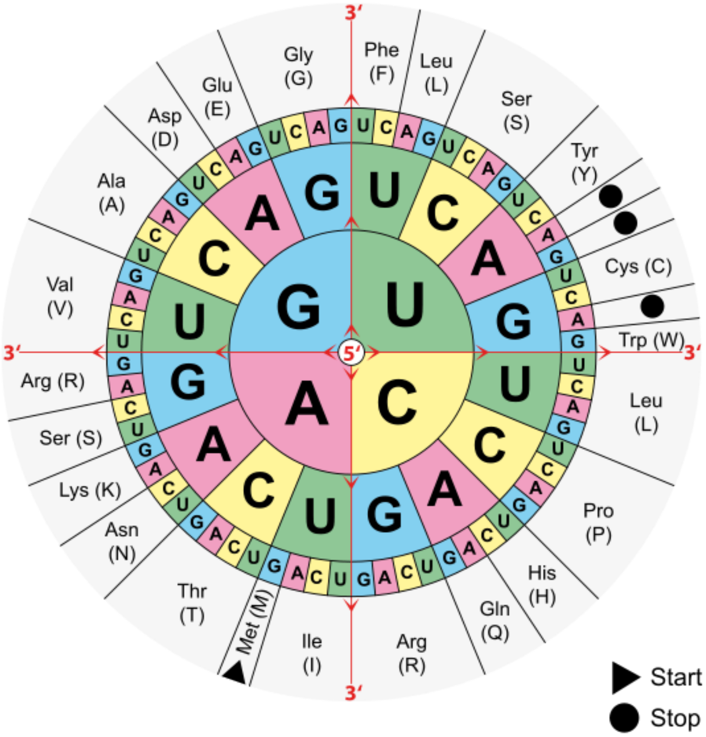
Commonly, an mRNA molecule will be translated simultaneously by several adjacent ribosomes. This increases the efficiency of protein synthesis. A single ribosome might translate an mRNA molecule in approximately one minute; so multiple ribosomes aboard a single transcript could produce multiple times the number of the same protein in the same minute. A **polyribosome** is a string of ribosomes translating a single mRNA strand.

Note:

Watch this video to learn about ribosomes. The ribosome binds to the mRNA molecule to start translation of its code into a protein. What happens to the small and large ribosomal subunits at the end of translation?

Translation works because each unique anticodon carries a specific amino acid. The same anticodon carries the same amino acid in virtually all living organisms on earth. This is why gene splicing and genetic engineering work. The complete collection of amino-acid/nucleotide-sequence relationships is called the **Genetic Code**. However, *the Genetic Code uses the complementary mRNA codon sequences rather than the tRNA anticodon sequences* to denote the relationship with the amino acid. The Genetic Code is presented in various formats. The two most common are the table format (not shown) and the wheel format shown below ([link]).

The Genetic Code



Genetic Code By Mouagip: Public domain, via Wikimedia Commons

The wheel is read from the center out (in the 5' to 3' direction). It is read by finding the first letter of the codon sequence in the inner most circle, then finding the second letter of the codon sequence from the four choices next to it in the second ring. The process is repeated in the third ring by choosing the third letter of the codon sequence from the four choices next to the second letter. The amino acid associated with the codon is found in the adjacent segment outside the circle.

Notice that four of the codon sequences have special significance. The sequence AUG is the start codon. This tells the ribosome where to start the translation. It also codes for the amino acid methionine. The other three special codon sequences are, UAA, UAG, and UGA. These are the stop codons. When the ribosome encounters these codons it stops the translation and releases the newly formed protein.

Looking at the amino acids around the outside of the wheel you also should notice that while each codon sequence is specific to a single amino acid, the reverse is not true. Most of the amino acids are associated with more than one codon. This *redundancy* in the Genetic Code can have a significant effect in some circumstances. Point mutations can occur in the DNA where one base is substituted for another. Even though the DNA has a mutation, the sequence now is different, if the resulting mRNA codon still codes for the same amino acid the resulting protein will not change. When this happens it is called a *silent mutation*.

Gene Mutations

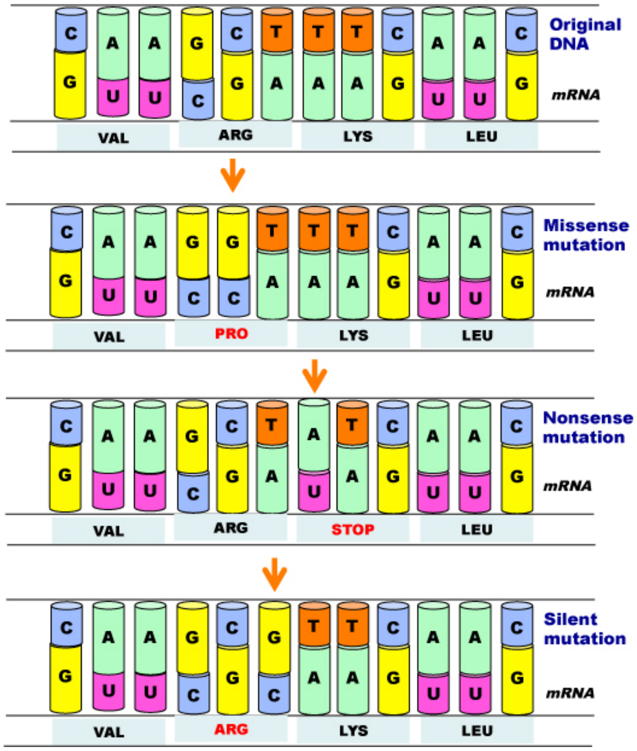
There are many ways the DNA can be altered by mutation. However, the changes to the original sequence are one of the three generic types:

- **Base Substitution:** where one base is replaced by another
- **Insertion:** where one or more bases are added to the sequence
- **Deletion:** where one or more bases are removed from the sequence

The effects of the mutation will differ depending on the specifics of each case. The effect generally will fall into one of four categories:

- **Missense:** where the mutation codes for a different amino acid at that location changing the resulting protein
- **Nonsense:** where the mutation codes for one of the three stop codons stopping the translation early
- **Silent:** where the DNA sequence changes but still codes for the same amino acid so the resulting protein is unchanged

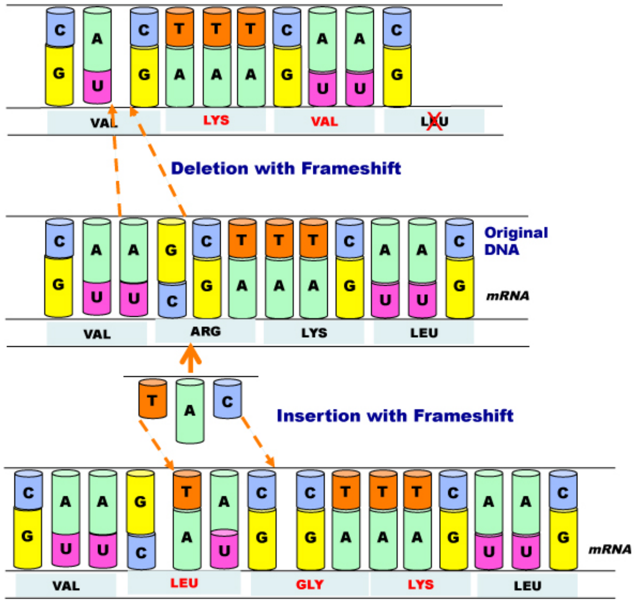
[link])



Effects From Single-Base Mutations

- **Frameshift:** where an insertion or deletion does not occur in a multiple of three bases or occurs in the middle of a triplet so it alters the reading frame and causes misreading of all the amino acids that follow

[link])



Examples of frameshift from insertions and deletions

The changes to the DNA can occur on a base-by-base basis. However, it also is common for some DNA segments to cause changes or mutations as a group. These are known as a transposable element or **Transposon**. This occurs both as *Chromosomal Rearrangement* on the same chromosome or as *Chromosomal Translocation* where the genetic material is exchanged between non-homologous chromosomes. Transposons tend to affect the coding in different ways than point mutations. They can affect gene expression in a number of ways.

A common gene element mutation is **Gene Duplication** where additional copies of a gene are inserted into the sequence during the replication process. The gene-repeats produced by this process can lead to additional expression of the gene which produces more of the protein in the body. Depending on the circumstance this may be beneficial or it may cause disease. It does not have to be the entire gene that duplicates. Sometimes only specific sequences within the gene are involved in the duplication. Huntington's disease is caused by a repeating section of the HTT gene. This section of the gene appears prone to duplication, at least in some individuals. When the number of repeats of this segment hits a threshold the Huntingtin protein it produces is altered (misshapen) enough that it no longer functions properly.

Gene duplication also plays an important role in evolution. From an evolutionary perspective there can be an advantage to generating additional copies of a functioning gene. The additional copies provide raw material for modification while leaving other unchanged copies of the gene to produce the original protein necessary to carry out the normal physiological processes. Without extra copies this would not be possible and every new protein would have to evolve from scratch rather than from modification of an existing functional protein.

Another way transposons can affect outcomes is by changing gene expression that is controlled by specific promoters and inhibitors. When a transposon relocates it can change the relationship between the genes it carries and their promoters or inhibitors. Depending on where the transposon moves it may make previously inactive genes active or active genes inactive. The transposon also may have an effect on surrounding genes. If it inserts between an active gene and its promoter it may inactivate that gene. Additionally, if a transposon inserts at an intron/exon splicing boundary it can change the splicing characteristics and alter the protein variants produced.

Chapter Review

DNA stores the information necessary for instructing the cell to perform all of its functions. Cells use the genetic code stored within DNA to build proteins, which ultimately determine the structure and function of the cell. This genetic code lies in the particular sequence of nucleotides that make up each gene along the DNA molecule. To “read” this code, the cell must perform two sequential steps. In the first step, *transcription*, the DNA code is converted into a RNA code. A molecule of messenger RNA that is complementary to a specific gene is synthesized in a process similar to DNA replication. The molecule of mRNA provides the code to synthesize a protein. In the process of *translation*, the mRNA attaches to a ribosome. Next, tRNA molecules shuttle the appropriate amino acids to the ribosome, one-by-one, coded by sequential triplet codons on the mRNA, until the protein is fully synthesized. When completed, the mRNA detaches from the ribosome, and the protein is released. Because the same codon is associated with the same amino acid in virtually all species a universal *Genetic Code* can be used to determine the amino acid sequence from any mRNA molecule. While each codon is associated with a single amino acid, the reverse is not true. Most amino acids have *redundancy*, meaning they are coded for by more than one codon. Typically, multiple ribosomes attach to a single mRNA molecule at once such that multiple proteins can be manufactured from the mRNA concurrently.

Interactive Link Questions

Exercise:

Problem:

Watch this video to learn about ribosomes. The ribosome binds to the mRNA molecule to start translation of its code into a protein. What happens to the small and large ribosomal subunits at the end of translation?

Solution:

They separate and move and are free to join translation of other segments of mRNA.

Review Questions

Exercise:

Problem:

Which of the following is *not* a difference between DNA and RNA?

a. DNA contains thymine whereas RNA contains uracil
b. DNA contains deoxyribose and RNA contains ribose

c. DNA contains alternating sugar-phosphate molecules whereas RNA does not contain sugars
d. RNA is single stranded and DNA is double stranded

Solution:

C

Exercise:

Problem:

Transcription and translation take place in the ________ and ________, respectively.

a. nucleus; cytoplasm
b. nucleolus; nucleus
c. nucleolus; cytoplasm
d. cytoplasm; nucleus

Solution:

A

Exercise:

Problem:

How many “letters” of an RNA molecule, in sequence, does it take to provide the code for a single amino acid?

a. 1
b. 2
c. 3
d. 4

Solution:

C

Exercise:

Problem: Which of the following is *not* made out of RNA?

a. the carriers that shuffle amino acids to a growing polypeptide strand
b. the ribosome
c. the messenger molecule that provides the code for protein synthesis
d. the intron

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

Briefly explain the similarities between transcription and DNA replication.

Solution:

Transcription and DNA replication both involve the synthesis of nucleic acids. These processes share many common features—particularly, the similar processes of initiation, elongation, and termination. In both cases the DNA molecule must be untwisted and separated, and the coding (i.e., sense) strand will be used as a template. Also, polymerases serve to add nucleotides to the growing DNA or mRNA strand. Both processes are signaled to terminate when completed.

Exercise:

Problem:

Contrast transcription and translation. Name at least three differences between the two processes.

Solution:

Transcription is really a “copy” process and translation is really an “interpretation” process, because transcription involves copying the DNA message into a very similar RNA message whereas translation involves converting the RNA message into the very different amino acid message. The two processes also differ in their location: transcription occurs in the nucleus and translation in the cytoplasm. The mechanisms by which the two processes are performed are also completely different: transcription utilizes polymerase enzymes to build mRNA whereas translation utilizes different kinds of RNA to build protein.

Exercise:

Problem:

Briefly explain what a silent mutation is and why it can happen.

Solution:

A silent mutation is a base substitution in the DNA sequence that, while different from the original, still codes for that same amino acid. This can happen because of the redundancy in the Genetic Code. While a given codon always codes for a specific amino acid, other codons may also code for that same amino acid. Codons that code for the same amino acid usually differ by only one base in their codon triplet. This means if a mutation happens at this location there is at least a one in three chance that it will also code for the same amino acid in the same place and there will be no change in the resulting protein.

Glossary

anticodon
consecutive sequence of three nucleotides on a tRNA molecule that is complementary to a specific codon on an mRNA molecule

codon
consecutive sequence of three nucleotides on an mRNA molecule that corresponds to a specific amino acid

exon
one of the coding regions of an mRNA molecule that remain after splicing

gene
functional length of DNA that provides the genetic information necessary to build a protein

gene expression
active interpretation of the information coded in a gene to produce a functional gene product

intron
non-coding regions of a pre-mRNA transcript that may be removed during splicing

messenger RNA (mRNA)
nucleotide molecule that serves as an intermediate in the genetic code between DNA and protein

polypeptide
chain of amino acids linked by peptide bonds

polyribosome
simultaneous translation of a single mRNA transcript by multiple ribosomes

promoter

region of DNA that signals transcription to begin at that site within the gene

proteome
full complement of proteins produced by a cell (determined by the cell’s specific gene expression)

ribosomal RNA (rRNA)
RNA that makes up the subunits of a ribosome

RNA polymerase
enzyme that unwinds DNA and then adds new nucleotides to a growing strand of RNA for the transcription phase of protein synthesis

spliceosome
complex of enzymes that serves to splice out the introns of a pre-mRNA transcript

splicing
the process of modifying a pre-mRNA transcript by removing certain, typically non-coding, regions

transcription
process of producing an mRNA molecule that is complementary to a particular gene of DNA

transfer RNA (tRNA)
molecules of RNA that serve to bring amino acids to a growing polypeptide strand and properly place them into the sequence

translation
process of producing a protein from the nucleotide sequence code of an mRNA transcript

triplet
consecutive sequence of three nucleotides on a DNA molecule that, when transcribed into an mRNA codon, corresponds to a particular amino acid

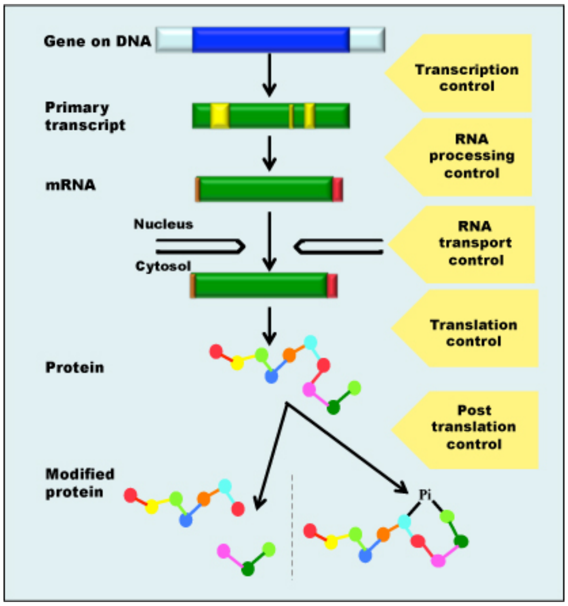
Gene Regulation and Expression

An overview of the multiple ways genes are regulated, including epigenetic, transcriptional, post-transcriptional, translational, and post-translational. This module combines material from existing OpenStax modules on these topics, specifically (m44536), (m44538), (m44539), and (m44542).

Life depends on the coordinated activities of proteins. They not only provide the foundation of three dimensional structures of living organisms, most enzymes that control the chemical activities of metabolism are proteins. The rate and duration of the millions of chemical reactions going on in your body depend on the correct presence and amounts of each protein enzyme. To further complicate the issue the same, or very similar, proteins often are used for different purposes in different parts of the body. The coordination of these activities is crucial and much of this class will focus on how these activities occur and are controlled in various body systems.

Because the proper expression and balance of proteins is so critical, living organisms have evolved the ability to control the production and expression of each protein at multiple levels in the gene-to-protein process. This provides the ability to tightly regulate the quantity, specificity, activity, and lifespan of a given protein. The combination allows cells to differentiate by silencing production of some proteins, regulating the production of other proteins, and modifying the activity of proteins already created. This section will cover the five locations where gene expression is most commonly controlled ([link]).

FIVE LEVELS OF GENE CONTROL



Source: Modified from: ArneLH, Wikimedia commons, Gene_expression_control.png

Chemical modifications occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. These changes can alter epigenetic accessibility, transcription, mRNA stability, or translation—all resulting in changes in expression of various genes. This is an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment. Because proteins are involved in every stage of gene regulation, the phosphorylation of a protein (depending on the protein that is modified) can alter accessibility to the chromosome, can alter translation (by altering transcription factor binding or function), can change nuclear shuttling (by influencing modifications to the nuclear pore complex), can alter RNA stability (by binding or not binding to the RNA to regulate its stability), can modify translation (increase or decrease), or can change post-translational modifications (add or remove phosphates or other chemical modifications).

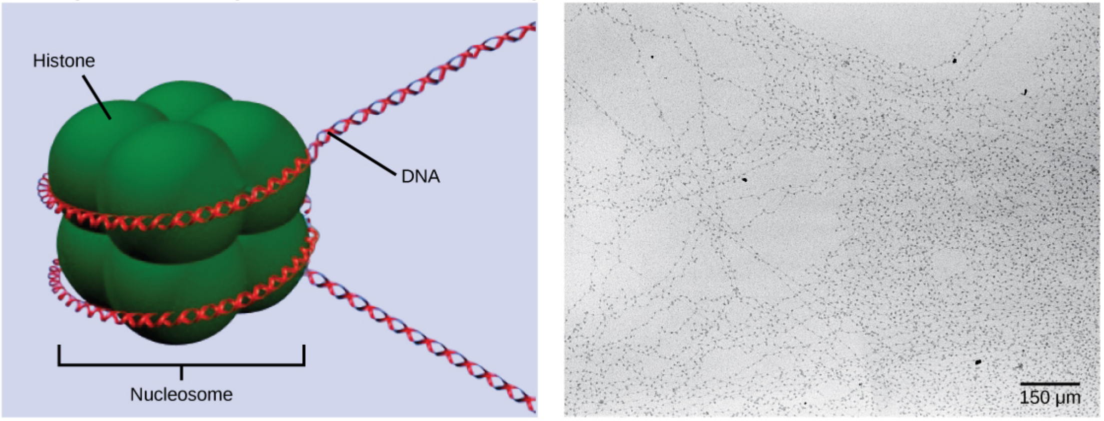
Transcription Control

Epigenetic Control: Regulating Access to Genes within the Chromosome

Eukaryotic gene expression begins with control of access to the DNA. This form of regulation, called epigenetic regulation, occurs even before transcription is initiated.

The human genome encodes over 20,000 genes; each of the 23 pairs of human chromosomes encodes thousands of genes. The DNA in the nucleus is precisely wound, folded, and compacted into chromosomes so that it will fit into the nucleus. It is also organized so that specific segments can be accessed as needed by a specific cell type.

The first level of organization, or packing, is the winding of DNA strands around histone proteins. Histones package and order DNA into structural units called nucleosome complexes, which can control the access of proteins to the DNA regions ([link]**a**). Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string ([link]**b**). These beads (histone proteins) can move along the string (DNA) and change the structure of the molecule.



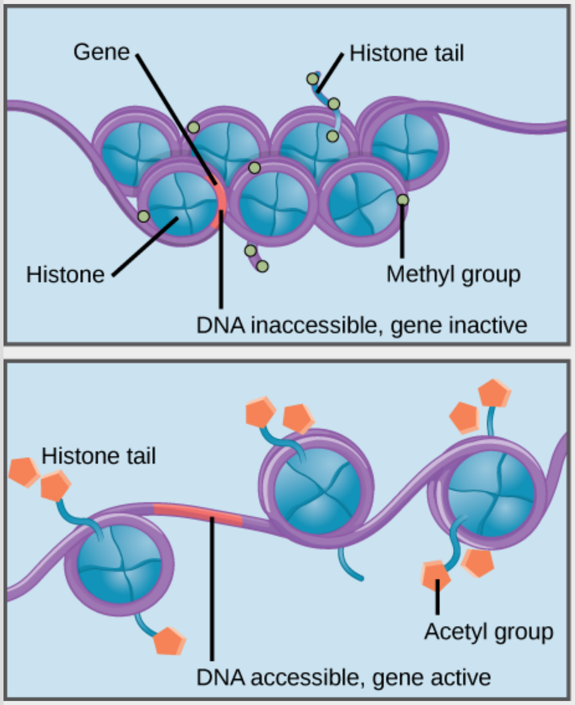
(a) (b)

DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit “micrograph”: modification of work by Chris Woodcock)

If DNA encoding a specific gene is to be transcribed into RNA, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow for the transcriptional machinery (RNA polymerase) to initiate transcription ([link]). Nucleosomes can move to open the chromosome structure to expose a segment of DNA, but do so in a very controlled manner.

Note:

Art Connection



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

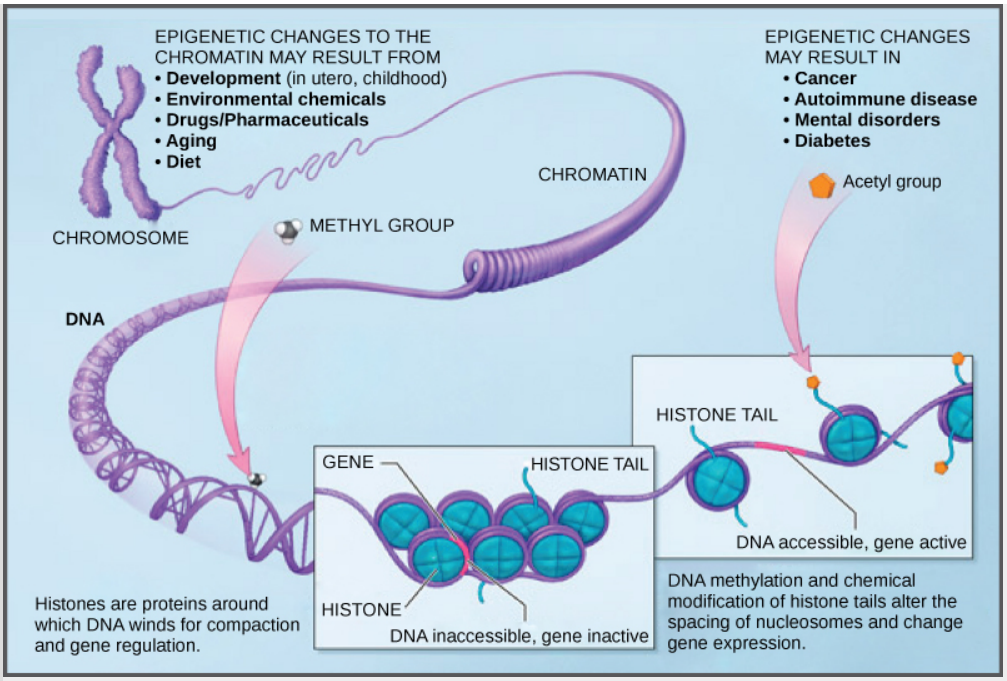
Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed.

Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

In females, one of the two X chromosomes is inactivated during embryonic development because of epigenetic changes to the chromatin. What impact do you think these changes would have on nucleosome packing?

How the histone proteins move is dependent on signals found on both the histone proteins and on the DNA. These signals are tags added to histone proteins and DNA that tell the histones if a chromosomal region should be open or closed ([link] depicts modifications to histone proteins and DNA). These tags are not permanent, but may be added or removed as needed. They are chemical modifications (phosphate, methyl, or acetyl groups) that are attached to specific amino acids in the protein or to the nucleotides of the DNA. The tags do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. When unmodified, the histone proteins have a large positive charge; by adding chemical modifications like acetyl groups, the charge becomes less positive.

The DNA molecule itself can also be modified. This occurs within very specific regions called CpG islands. These are stretches with a high frequency of cytosine and guanine dinucleotide DNA pairs (CG) found in the promoter regions of genes. When this configuration exists, the cytosine member of the pair can be methylated (a methyl group is added). This modification changes how the DNA interacts with proteins, including the histone proteins that control access to the region. Highly methylated (hypermethylated) DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive.


EPIGENETIC CHANGES TO THE CHROMATIN MAY RESULT FROM
• Development (in utero, childhood)
• Environmental chemicals
• Drugs/Pharmaceuticals
• Aging
• Diet
EPIGENETIC CHANGES MAY RESULT IN
• Cancer
• Autoimmune disease
• Mental disorders
• Diabetes
CHROMATIN
Acetyl group
CHROMOSOME
METHYL GROUP
DNA
HISTONE TAIL
GENE
HISTONE TAIL
DNA accessible, gene active
DNA methylation and chemical modification of histone tails alter the spacing of nucleosomes and change gene expression.
Histones are proteins around which DNA winds for compaction and gene regulation.
HISTONE
DNA inaccessible, gene inactive

Histone proteins and DNA nucleotides can be modified chemically. Modifications affect nucleosome spacing and gene expression. (credit: modification of work by NIH)

This type of gene regulation is called epigenetic regulation. Epigenetic means “around genetics.” The changes that occur to the histone proteins and DNA do not alter the nucleotide sequence and are not permanent. Instead, these changes are temporary (although they often persist through multiple rounds of cell division) and alter the chromosomal structure (open or closed) as needed. A gene can be turned on or off depending upon the location and modifications to the histone proteins and DNA. If a gene is to be transcribed, the histone proteins and DNA are modified surrounding the chromosomal region encoding that gene. This opens the chromosomal region to allow access for RNA polymerase and other proteins, called **transcription factors**, to bind to the promoter region, located just upstream of the gene, and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have different modifications that signal a closed chromosomal configuration. In this closed configuration, the RNA polymerase and transcription factors do not have access to the DNA and transcription cannot occur ([link]).

Note:
Link to Learning

View this video that describes how epigenetic regulation controls gene expression.

Regulation of Transcription

The transcription of genes in eukaryotes requires the actions of an RNA polymerase to bind to a sequence upstream of a gene to initiate transcription. Unlike prokaryotic cells, the eukaryotic RNA polymerase requires other proteins, or transcription factors, to facilitate transcription initiation. **Transcription factors** are proteins that bind to the promoter sequence and other regulatory sequences to control the transcription of the target gene. RNA polymerase by itself cannot initiate transcription in eukaryotic cells. Transcription factors must bind to the promoter region first and recruit RNA polymerase to the site for transcription to be established.

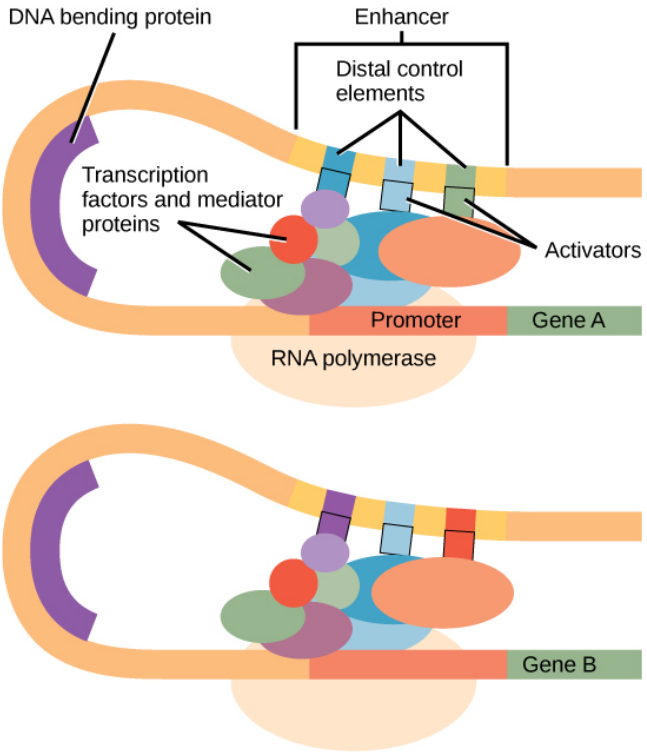
Note:
Link to Learning

View the process of transcription—the making of RNA from a DNA template—at this site.

The Promoter and the Transcription Machinery

Genes are organized to make the control of gene expression easier. The promoter region is immediately upstream of the coding sequence. This region can be short (only a few nucleotides in length) or quite long (hundreds of nucleotides long). The longer the promoter, the more available space for proteins to bind. This also adds more control to the transcription process. The length of the promoter is gene-specific and can differ dramatically between genes. Consequently, the level of control of gene expression can also differ quite dramatically between genes. The purpose of the promoter is to bind transcription factors that control the initiation of transcription.

Within the promoter region, just upstream of the transcriptional start site, resides the TATA box. This box is simply a repeat of thymine and adenine dinucleotides (literally, TATA repeats). RNA polymerase binds to the transcription initiation complex, allowing transcription to occur. To initiate transcription, a transcription factor (TFIID) is the first to bind to the TATA box. Binding of TFIID recruits other transcription factors, including TFIIB, TFIIE, TFIIF, and TFIIH to the TATA box. Once this complex is assembled, RNA polymerase can bind to its upstream sequence. When bound along with the transcription factors, RNA polymerase is phosphorylated. This releases part of the protein from the DNA to activate the transcription initiation complex and places RNA polymerase in the correct orientation to begin transcription; DNA-bending protein brings the enhancer, which can be quite a distance from the gene, in contact with transcription factors and mediator proteins ([link]).



An enhancer is a DNA sequence that promotes transcription. Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors. Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.

In addition to the general transcription factors, other transcription factors can bind to the promoter to regulate gene transcription. These transcription factors bind to the promoters of a specific set of genes. They are not general transcription factors that bind to every promoter complex, but are recruited to a specific sequence on the promoter of a specific gene. There are hundreds of transcription factors in a cell that each bind specifically to a particular DNA sequence motif. When transcription factors bind to the promoter just upstream of the encoded gene, it is referred to as a ***cis*-acting element**, because it is on the same chromosome just next to the gene. The region that a particular transcription factor binds to is called the **transcription factor binding site**. Transcription factors respond to environmental stimuli that cause the proteins to find their binding sites and initiate transcription of the gene that is needed.

Enhancers and Transcription

In some eukaryotic genes, there are regions that help increase or enhance transcription. These regions, called **enhancers**, are not necessarily close to the genes they enhance. They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away.

Enhancer regions are binding sequences, or sites, for transcription factors. When a DNA-bending protein binds, the shape of the DNA changes ([link]). This shape change allows for the interaction of the activators bound to the enhancers with the transcription factors bound to the promoter region and the RNA polymerase. Whereas DNA is generally depicted as a straight line in two dimensions, it is actually a three-dimensional object. Therefore, a nucleotide sequence thousands of nucleotides away can fold over and interact with a specific promoter.

Turning Genes Off: Transcriptional Repressors

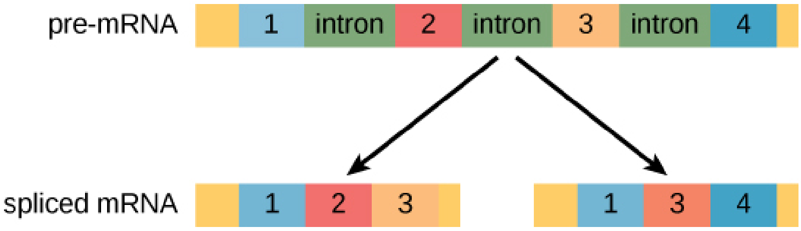
Eukaryotic cells also have mechanisms to prevent transcription. Transcriptional repressors can bind to promoter or enhancer regions and block transcription. Like the transcriptional activators, repressors respond to external stimuli to prevent the binding of activating transcription factors.Insert paragraph text here.

RNA-Processing Control

RNA is transcribed, but must be processed into a mature form before translation can begin. This processing after an RNA molecule has been transcribed, but before it is translated into a protein, is called post-transcriptional modification. As with the epigenetic and transcriptional stages of processing, this post-transcriptional step can also be regulated to control gene expression in the cell. If the RNA is not processed, shuttled, or translated, then no protein will be synthesized.

RNA splicing, the first stage of post-transcriptional control

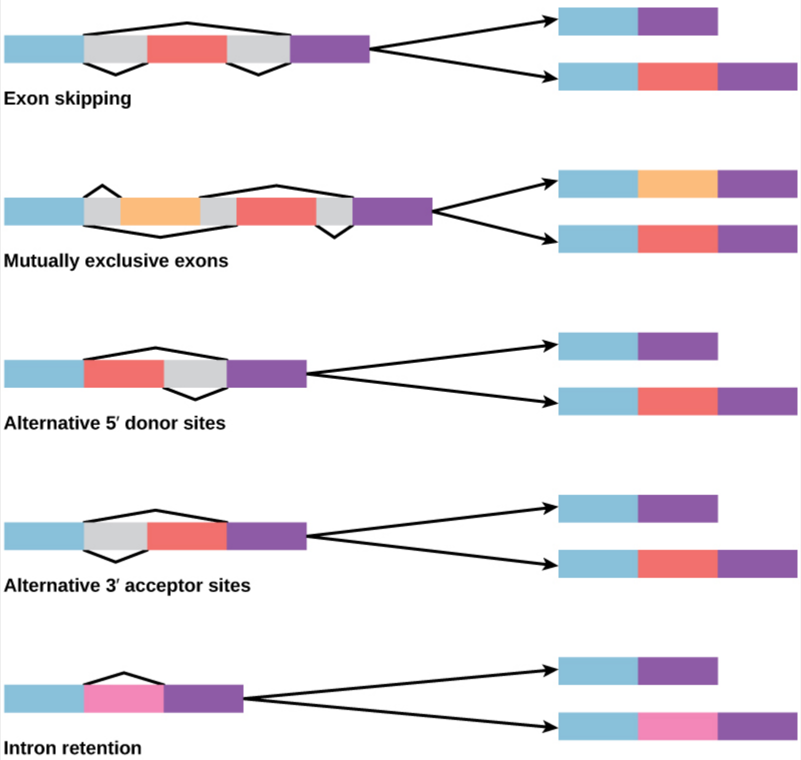
In eukaryotic cells, the RNA transcript often contains regions, called introns, that are removed prior to translation. The regions of RNA that code for protein are called exons ([link]). After an RNA molecule has been transcribed, but prior to its departure from the nucleus to be translated, the RNA is processed and the introns are removed by splicing.



Pre-mRNA can be alternatively spliced to create different proteins.

Note:
Evolution Connection
Alternative RNA Splicing
In the 1970s, genes were first observed that exhibited alternative RNA splicing. Alternative RNA splicing is a mechanism that allows different protein products to be produced from one gene when different

combinations of introns, and sometimes exons, are removed from the transcript ([link]). This alternative splicing can be haphazard, but more often it is controlled and acts as a mechanism of gene regulation, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells or at different stages of development. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; according to one estimate, 70 percent of genes in humans are expressed as multiple proteins through alternative splicing.



There are five basic modes of alternative splicing.

How could alternative splicing evolve? Introns have a beginning and ending recognition sequence; it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and instead find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such intron skipping, but mutations are likely to lead to their failure. Such “mistakes” would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is alternative splicing rather than mutations in a sequence. However, alternative splicing would create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way by providing genes that may evolve without eliminating the original, functional protein.

Note:
Link to Learning

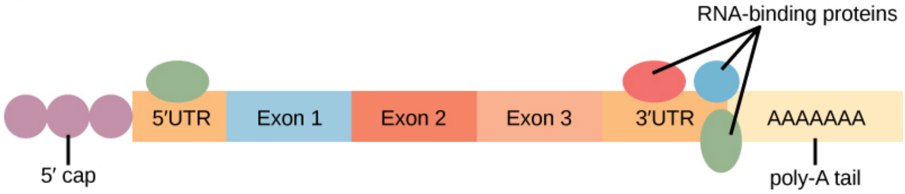
Visualize how mRNA splicing happens by watching the process in action in this video.

Control of RNA Stability

Before the mRNA leaves the nucleus, it is given two protective "caps" that prevent the end of the strand from degrading during its journey. The **5' cap**, which is placed on the 5' end of the mRNA, is usually composed of a methylated guanosine triphosphate molecule (GTP). The **poly-A tail**, which is attached to the 3' end, is usually composed of a series of adenine nucleotides. Once the RNA is transported to the cytoplasm, the length of time that the RNA resides there can be controlled. Each RNA molecule has a defined lifespan and decays at a specific rate. This rate of decay can influence how much protein is in the cell. If the decay rate is increased, the

RNA will not exist in the cytoplasm as long, shortening the time for translation to occur. Conversely, if the rate of decay is decreased, the RNA molecule will reside in the cytoplasm longer and more protein can be translated. This rate of decay is referred to as the RNA stability. If the RNA is stable, it will be detected for longer periods of time in the cytoplasm.

Binding of proteins to the RNA can influence its stability. Proteins, called **RNA-binding proteins**, or RBPs, can bind to the regions of the RNA just upstream or downstream of the protein-coding region. These regions in the RNA that are not translated into protein are called the **untranslated regions**, or UTRs. They are not introns (those have been removed in the nucleus). Rather, these are regions that regulate mRNA localization, stability, and protein translation. The region just before the protein-coding region is called the **5' UTR**, whereas the region after the coding region is called the **3' UTR** ([link]). The binding of RBPs to these regions can increase or decrease the stability of an RNA molecule, depending on the specific RBP that binds.



The protein-coding region of mRNA is flanked by 5' and 3' untranslated regions (UTRs). The presence of RNA-binding proteins at the 5' or 3' UTR influences the stability of the RNA molecule.

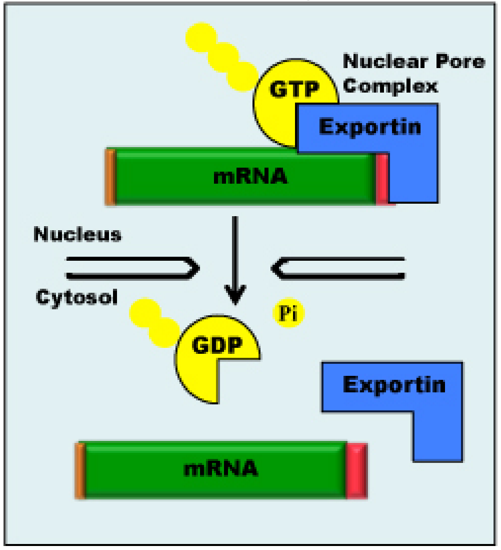
RNA Stability and microRNAs

In addition to RBPs that bind to and control (increase or decrease) RNA stability, other elements called microRNAs can bind to the RNA molecule. These **microRNAs**, or miRNAs, are short RNA molecules that are only 21–24 nucleotides in length. The miRNAs are made in the nucleus as longer pre-miRNAs. These pre-miRNAs are chopped into mature miRNAs by a protein called **dicer**. Like transcription factors and RBPs, mature miRNAs recognize a specific sequence and bind to the RNA; however, miRNAs also associate with a ribonucleoprotein complex called the **RNA-induced silencing complex (RISC)**. RISC binds along with the miRNA to degrade the target mRNA. Together, miRNAs and the RISC complex rapidly destroy the RNA molecule.

RNA-Transport Control

Gene expression requires the movement of mRNA molecules from the nucleus to the cytoplasm. The export process is controlled by **Nuclear Pore Complexes (NPCs)**, a combination of the RNA molecule and molecules generically called **exportins**. Although the exact process is not well understood for mRNA the generic process appears similar to export of other kinds of RNA and protein molecules.

The Nuclear Pore Complex binds with a **guanosine triphosphate (GTP)** molecule that is hydrolyzed by a **GTPase** molecule into GDP. The energy released by cleaving the phosphate bond is used to shuttle the mRNA through the nuclear pore. Once through the pore the the Nuclear Pore Complex dissociates and the GDP molecule returns to the nucleus leaving the mRNA in the cytoplasm ([link]).



A Nuclear Pore Complex

(mRNA, Exportin, and GTP) transport the mRNA through the nuclear pore. In the process one of the phosphate molecules dissociates. Once through the nuclear pore the complex breaks apart into separate parts with the mRNA now free in the cytoplasm.

The physical and electrical charge characteristics of the nuclear pore interact with the Nuclear Pore Complex to selectively move substances through the pore. The availability of the appropriate receptors and binding proteins regulate how much mRNA can pass though the pores and therefore amount of protein that can be translated.

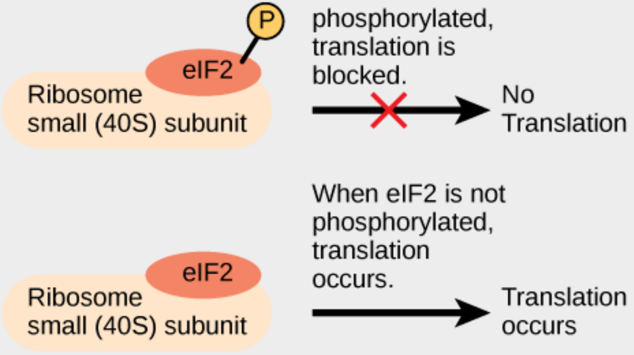
Translation Control

After the RNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

The Initiation Complex and Translation Rate

Like transcription, translation is controlled by proteins that bind and initiate the process. In translation, the complex that assembles to start the process is referred to as the **initiation complex**. The first protein to bind to the RNA to initiate translation is the **eukaryotic initiation factor-2 (eIF-2)**. The eIF-2 protein is active when it binds to the high-energy molecule **guanosine triphosphate (GTP)**. GTP provides the energy to start the reaction by giving up a phosphate and becoming **guanosine diphosphate (GDP)**. The eIF-2 protein bound to GTP binds to the small **40S ribosomal subunit**. When bound, the methionine initiator tRNA associates with the eIF-2/40S ribosome complex, bringing along with it the mRNA to be translated. At this point, when the initiator complex is assembled, the GTP is converted into GDP and energy is released. The phosphate and the eIF-2 protein are released from the complex and the large **60S ribosomal subunit** binds to translate the RNA. The binding of eIF-2 to the RNA is controlled by phosphorylation. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded ([link]). When eIF-2 remains unphosphorylated, it binds the RNA and actively translates the protein.

Note:
Art Connection


P
eIF2
Ribosome small (40S) subunit
When eIF2 is phosphorylated, translation is blocked.
No Translation
eIF2
Ribosome small (40S) subunit
When eIF2 is not phosphorylated, translation occurs.
Translation occurs

Gene expression can be controlled by factors that bind the translation initiation complex.

An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer’s, Parkinson’s, and Huntington’s. What impact do you think this might have on protein synthesis?

Post-Translation Control

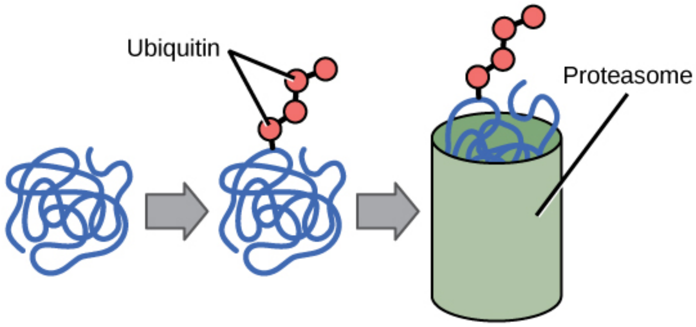
Three major mechanisms for post-translational modification are:

- Cleaving the protein (removing amino acids)
- Adding chemical groups to the protein
- Combining the protein with one or more other proteins

Some proteins begin in their inactive form. For example angiotensinogen, a protein that is created in the liver and circulates in the blood must first be cleaved by the peptide renin (from the kidneys) and then have additional enzymatic modification before it becomes active. Other proteins, like acetylcholine are created in their active form and stored in vesicles until they are needed. Once released the protein is quickly cleaved by acetylcholinesterase to make it inactive.

Some proteins are chemically modified with the addition of chemical groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins regulates their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, the cytoplasm, or attached to the plasma membrane. Adding and removing phosphate groups to change protein conformations is a major way cells use ATP to power cellular functions. Examples include the cocking and release of myosin heads during muscle contraction and the opening and closing of sodium/potassium pumps in the cell membrane.

The addition of an ubiquitin group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein lifespan is complete. These proteins are moved to the **proteasome**, an organelle that functions to remove proteins, to be degraded ([link]). One way to control gene expression, therefore, is to alter the longevity of the protein.



Proteins with ubiquitin tags are marked for degradation within the proteasome.

Section Summary

In eukaryotic cells, the first stage of gene expression control occurs at the epigenetic level. Epigenetic mechanisms control access to the chromosomal region to allow genes to be turned on or off. These mechanisms control how DNA is packed into the nucleus by regulating how tightly the DNA is wound around histone proteins. The addition or removal of chemical modifications (or flags) to histone proteins or DNA signals to the cell to open or close a chromosomal region. Therefore, eukaryotic cells can control whether a gene is expressed by controlling accessibility to transcription factors and the binding of RNA polymerase to initiate transcription.

To start transcription, general transcription factors, such as TFIID, TFIIH, and others, must first bind to the TATA box and recruit RNA polymerase to that location. The binding of additional regulatory transcription factors to *cis*-acting elements will either increase or prevent transcription. In addition to promoter sequences, enhancer regions help augment transcription. Enhancers can be upstream, downstream, within a gene itself, or on other chromosomes. Transcription factors bind to enhancer regions to increase or prevent transcription.

Post-transcriptional control can occur at any stage after transcription, including RNA splicing, nuclear shuttling, and RNA stability. Once RNA is transcribed, it must be processed to create a mature RNA that is ready to be translated. This involves the removal of introns that do not code for protein. Spliceosomes bind to the signals that mark the exon/intron border to remove the introns and ligate the exons together. Once this occurs, the RNA is mature and can be translated. RNA is created and spliced in the nucleus, but needs to be transported to the cytoplasm to be translated. RNA is transported to the cytoplasm through the nuclear pore complex. Once the RNA is in the cytoplasm, the length of time it resides there before being degraded, called RNA stability, can also be altered to control the overall amount of protein that is synthesized. The RNA stability can be increased, leading to longer residency time in the cytoplasm, or decreased, leading to shortened time and less protein synthesis. RNA stability is controlled by RNA-binding proteins (RPBs) and microRNAs (miRNAs). These RPBs and miRNAs bind to the 5' UTR or the 3' UTR of the RNA to increase or decrease RNA stability. Depending on the RBP, the stability can be increased or decreased significantly; however, miRNAs always decrease stability and promote decay.

Changing the status of the RNA or the protein itself can affect the amount of protein, the function of the protein, or how long it is found in the cell. To translate the protein, a protein initiator complex must assemble on the RNA. Modifications (such as phosphorylation) of proteins in this complex can prevent proper translation from occurring. Once a protein has been synthesized, it can be modified (phosphorylated, acetylated, methylated, or ubiquitinated). These post-translational modifications can greatly impact the stability, degradation, or function of the protein.

Art Connections

Exercise:

Problem:

[link] In females, one of the two X chromosomes is inactivated during embryonic development because of epigenetic changes to the chromatin. What impact do you think these changes would have on nucleosome packing?

Solution:

[link] The nucleosomes would pack more tightly together.

Exercise:

Problem:

[link] An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer’s, Parkinson’s, and Huntington’s. What impact do you think this might have on protein synthesis?

Solution:

[link] Protein synthesis would be inhibited.

Review Questions

Exercise:

Problem: What are epigenetic modifications?

a. the addition of reversible changes to histone proteins and DNA
b. the removal of nucleosomes from the DNA
c. the addition of more nucleosomes to the DNA
d. mutation of the DNA sequence

Solution:

A

Exercise:

Problem: Which of the following are true of epigenetic changes?

a. allow DNA to be transcribed

b. move histones to open or close a chromosomal region
c. are temporary
d. all of the above

Solution:

D

Exercise:

Problem:

The binding of _________ is required for transcription to start.

a. a protein
b. DNA polymerase
c. RNA polymerase
d. a transcription factor

Solution:

C

Exercise:

Problem:

What will result from the binding of a transcription factor to an enhancer region?

a. decreased transcription of an adjacent gene
b. increased transcription of a distant gene
c. alteration of the translation of an adjacent gene
d. initiation of the recruitment of RNA polymerase

Solution:

B

Exercise:

Problem:

Which of the following are involved in post-transcriptional control?

1. control of RNA splicing
2. control of RNA shuttling
3. control of RNA stability
4. all of the above

Solution:

D

Exercise:

Problem:

Binding of an RNA binding protein will _________ the stability of the RNA molecule.

a. increase
b. decrease
c. neither increase nor decrease
d. either increase or decrease

Solution:

D

Exercise:

Problem:

Post-translational modifications of proteins can affect which of the following?

a. protein function

b. transcriptional regulation
c. chromatin modification
d. all of the above

Solution:

A

Free Response

Exercise:

Problem:

In cancer cells, alteration to epigenetic modifications turns off genes that are normally expressed. Hypothetically, how could you reverse this process to turn these genes back on?

Solution:

You can create medications that reverse the epigenetic processes (to add histone acetylation marks or to remove DNA methylation) and create an open chromosomal configuration.

Exercise:

Problem:

A mutation within the promoter region can alter transcription of a gene. Describe how this can happen.

Solution:

A mutation in the promoter region can change the binding site for a transcription factor that normally binds to increase transcription. The mutation could either decrease the ability of the transcription factor to bind, thereby decreasing transcription, or it can increase the ability of the transcription factor to bind, thus increasing transcription.

Exercise:

Problem:

What could happen if a cell had too much of an activating transcription factor present?

Solution:

If too much of an activating transcription factor were present, then transcription would be increased in the cell. This could lead to dramatic alterations in cell function.

Exercise:

Problem:

Describe how RBPs can prevent miRNAs from degrading an RNA molecule.

Solution:

RNA binding proteins (RBP) bind to the RNA and can either increase or decrease the stability of the RNA. If they increase the stability of the RNA molecule, the RNA will remain intact in the cell for a longer period of time than normal. Since both RBPs and miRNAs bind to the RNA molecule, RBP can potentially bind first to the RNA and prevent the binding of the miRNA that will degrade it.

Exercise:

Problem:

How can external stimuli alter post-transcriptional control of gene expression?

Solution:

External stimuli can modify RNA-binding proteins (i.e., through phosphorylation of proteins) to alter their activity.

Exercise:

Problem:

Protein modification can alter gene expression in many ways. Describe how phosphorylation of proteins can alter gene expression.

Solution:

Because proteins are involved in every stage of gene regulation, phosphorylation of a protein (depending on the protein that is modified) can alter accessibility to the chromosome, can alter translation (by altering the transcription factor binding or function), can change nuclear shuttling (by influencing modifications to the nuclear pore complex), can alter RNA stability (by binding or not binding to the RNA to regulate its stability), can modify translation (increase or decrease), or can change post-translational modifications (add or remove phosphates or other chemical modifications).

Exercise:

Problem:

Alternative forms of a protein can be beneficial or harmful to a cell. What do you think would happen if too much of an alternative protein bound to the 3' UTR of an RNA and caused it to degrade?

Solution:

If the RNA degraded, then less of the protein that the RNA encodes would be translated. This could have dramatic implications for the cell.

Exercise:

Problem:

Changes in epigenetic modifications alter the accessibility and transcription of DNA. Describe how environmental stimuli, such as ultraviolet light exposure, could modify gene expression.

Solution:

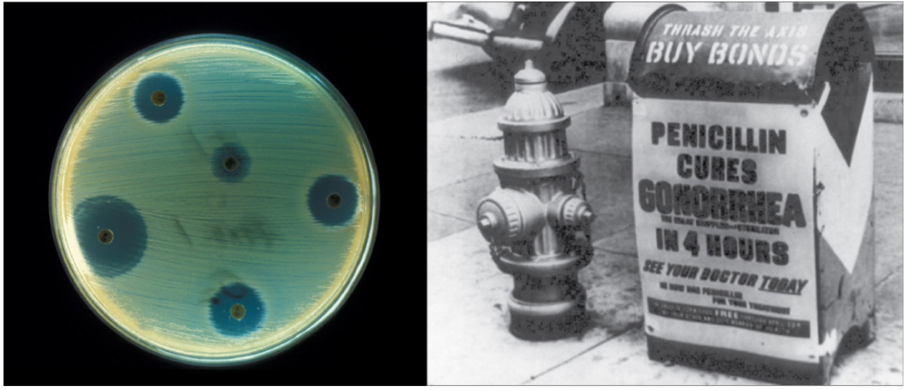
Environmental stimuli, like ultraviolet light exposure, can alter the modifications to the histone proteins or DNA. Such stimuli may change an actively transcribed gene into a silenced gene by removing acetyl groups from histone proteins or by adding methyl groups to DNA.

Biotechnology

By the end of this section, you will be able to:

- Describe gel electrophoresis
- Explain molecular and reproductive cloning
- Describe uses of biotechnology in medicine and agriculture

Biotechnology is the use of biological agents for technological advancement. Biotechnology was used for breeding livestock and crops long before the scientific basis of these techniques was understood. Since the discovery of the structure of DNA in 1953, the field of biotechnology has grown rapidly through both academic research and private companies. The primary applications of this technology are in medicine (production of vaccines and antibiotics) and agriculture (genetic modification of crops, such as to increase yields). Biotechnology also has many industrial applications, such as fermentation, the treatment of oil spills, and the production of biofuels ([link]).



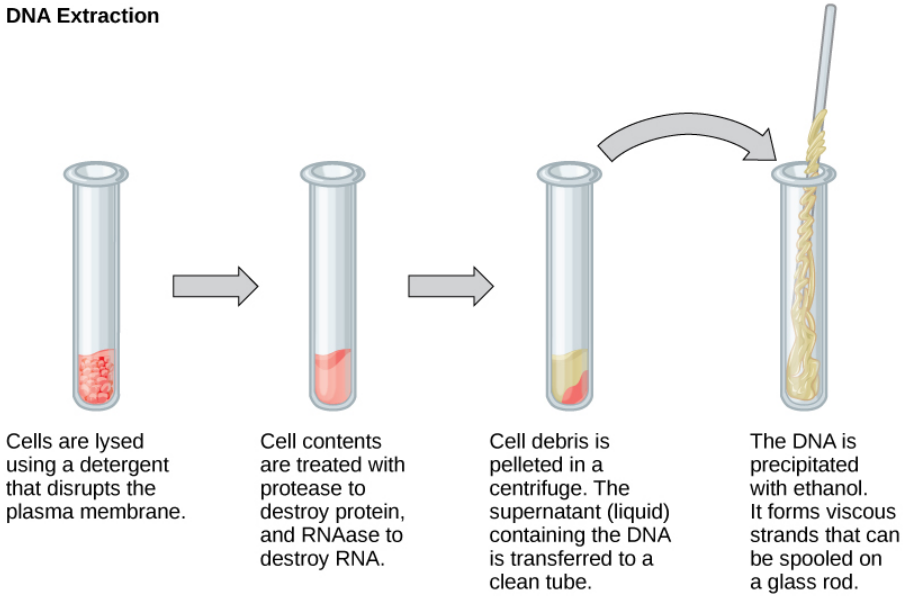
Antibiotics are chemicals produced by fungi, bacteria, and other organisms that have antimicrobial properties. The first antibiotic discovered was penicillin. Antibiotics are now commercially produced and tested for their potential to inhibit bacterial growth. (credit "advertisement": modification of work by NIH; credit "test plate": modification of work by Don Stalons/CDC; scale-bar data from Matt Russell)

Basic Techniques to Manipulate Genetic Material (DNA and RNA)

To understand the basic techniques used to work with nucleic acids, remember that nucleic acids are macromolecules made of nucleotides (a sugar, a phosphate, and a nitrogenous base) linked by phosphodiester bonds. The phosphate groups on these molecules each have a net negative charge. An entire set of DNA molecules in the nucleus is called the genome. DNA has two complementary strands linked by hydrogen bonds between the paired bases. The two strands can be separated by exposure to high temperatures (DNA denaturation) and can be reannealed by cooling. The DNA can be replicated by the DNA polymerase enzyme. Unlike DNA, which is located in the nucleus of eukaryotic cells, RNA molecules leave the nucleus. The most common type of RNA that is analyzed is the messenger RNA (mRNA) because it represents the protein-coding genes that are actively expressed. However, RNA molecules present some other challenges to analysis, as they are often less stable than DNA.

DNA and RNA Extraction

To study or manipulate nucleic acids, the DNA or RNA must first be isolated or extracted from the cells. Various techniques are used to extract different types of DNA ([link]). Most nucleic acid extraction techniques involve steps to break open the cell and use enzymatic reactions to destroy all macromolecules that are not desired (such as degradation of unwanted molecules and separation from the DNA sample). Cells are broken using a **lysis buffer** (a solution which is mostly a detergent); lysis means “to split.” These enzymes break apart lipid molecules in the cell membranes and nuclear membranes. Macromolecules are inactivated using enzymes such as **proteases** that break down proteins, and **ribonucleases** (RNAses) that break down RNA. The DNA is then precipitated using alcohol. Human genomic DNA is usually visible as a gelatinous, white mass. The DNA samples can be stored frozen at –80°C for several years.


DNA Extraction
Cells are lysed using a detergent that disrupts the plasma membrane.
Cell contents are treated with protease to destroy protein, and RNAase to destroy RNA.
Cell debris is pelleted in a centrifuge. The supernatant (liquid) containing the DNA is transferred to a clean tube.
The DNA is precipitated with ethanol. It forms viscous strands that can be spooled on a glass rod.

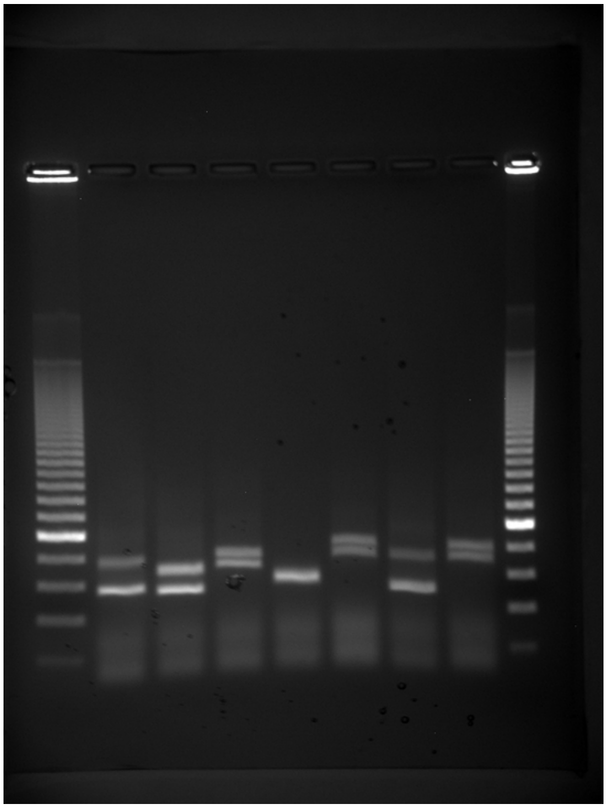
This diagram shows the basic method used for extraction of DNA.

DNA molecules are very long and contain many genes. The current processes for sequencing DNA only work on segments that are 1,000 base pairs or smaller. Therefore the DNA molecules must be cut (digested) into smaller pieces at known sequence locations before they can be further analyzed or used. This is done using **Restriction Enzymes**. Restriction enzymes were first discovered in bacteria, and more than 3,000 have been identified and studied. They fall into five types that vary in terms of how close to the recognition sequence (restriction site) they cut the DNA and whether they require ATP to carry out the restriction process. Restriction enzymes are used in many applications where it is important to cut DNA at specific locations. This includes both the extraction and insertion of genes from one DNA strand to another in gene translocation, identifying how many copies of a specific gene are present in a specific individual, comparing differences in fragment length between individuals (DNA fingerprinting), and potentially deleting or fixing faulty sequences in individuals with genetic disorders.

RNA analysis is performed to study gene expression patterns in cells. RNA is naturally very unstable because RNAses are commonly present in nature and very difficult to inactivate. Similar to DNA, RNA extraction involves the use of various buffers and enzymes to inactivate macromolecules and preserve the RNA.

Gel Electrophoresis

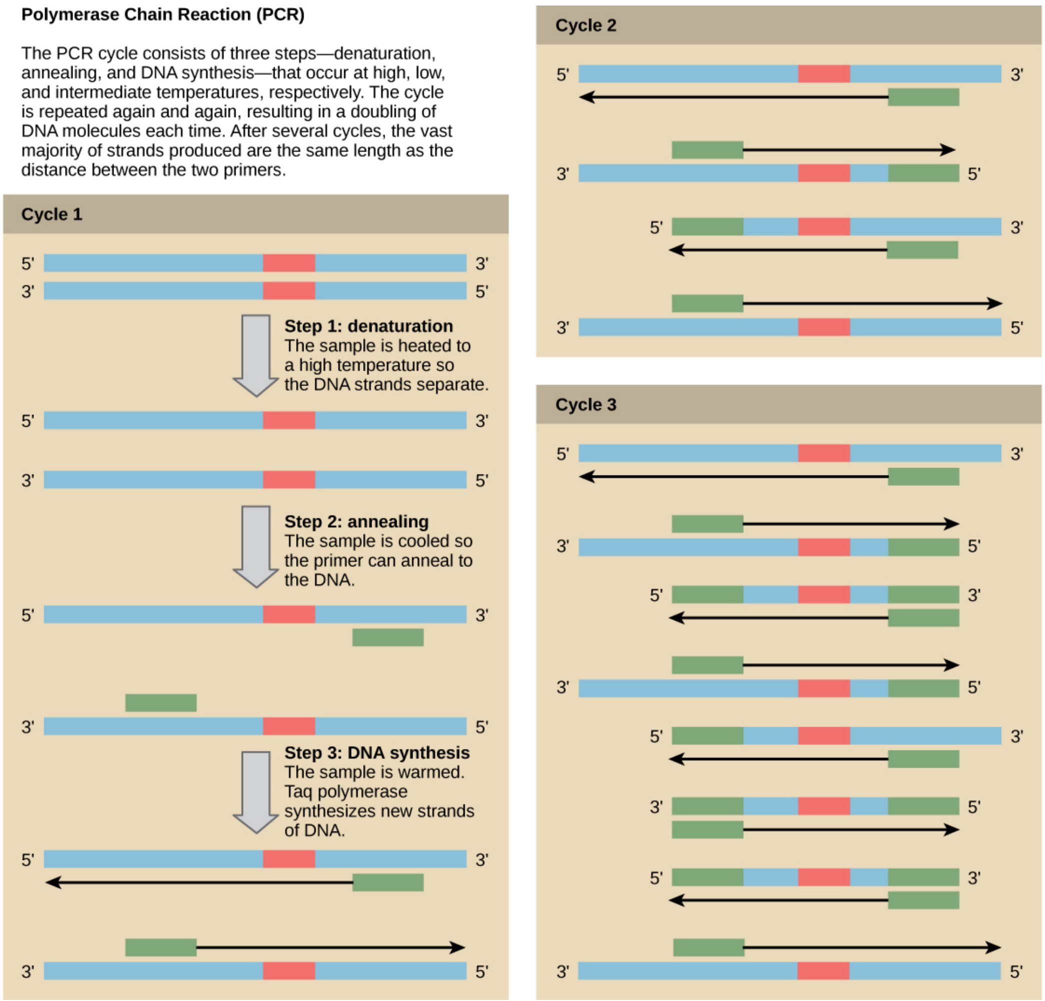
Because nucleic acids are negatively charged ions at neutral or basic pH in an aqueous environment, they can be mobilized by an electric field. **Gel electrophoresis** is a technique used to separate molecules on the basis of size, using this charge. The nucleic acids can be separated as whole chromosomes or fragments. The nucleic acids are loaded into a slot near the negative electrode of a semisolid, porous gel matrix and pulled toward the positive electrode at the opposite end of the gel. Smaller molecules move through the pores in the gel faster than larger molecules; this difference in the rate of migration separates the fragments on the basis of size. There are molecular weight standard samples that can be run alongside the molecules to provide a size comparison. Nucleic acids in a gel matrix can be observed using various fluorescent or colored dyes. Distinct nucleic acid fragments appear as bands at specific distances from the top of the gel (the negative electrode end) on the basis of their size ([link]). A mixture of genomic DNA fragments of varying sizes appear as a long smear, whereas uncut genomic DNA is usually too large to run through the gel and forms a single large band at the top of the gel.



Shown are DNA fragments from seven samples run on a gel, stained with a fluorescent dye, and viewed under UV light. (credit: James Jacob, Tompkins Cortland Community College)

Amplification of Nucleic Acid Fragments by Polymerase Chain Reaction

Although genomic DNA is visible to the naked eye when it is extracted in bulk, DNA analysis often requires focusing on one or more specific regions of the genome. **Polymerase chain reaction** (**PCR**) is a technique used to amplify specific regions of DNA for further analysis ([link]). PCR is used for many purposes in laboratories, such as the cloning of gene fragments to analyze genetic diseases, identification of contaminant foreign DNA in a sample, and the amplification of DNA for sequencing. More practical applications include the determination of paternity and detection of genetic diseases.



Polymerase Chain Reaction (PCR)

The PCR cycle consists of three steps—denaturation, annealing, and DNA synthesis—that occur at high, low, and intermediate temperatures, respectively. The cycle is repeated again and again, resulting in a doubling of DNA molecules each time. After several cycles, the vast majority of strands produced are the same length as the distance between the two primers.

Step 1: denaturation The sample is heated to a high temperature so the DNA strands separate.

Step 2: annealing The sample is cooled so the primer can anneal to the DNA.

Step 3: DNA synthesis The sample is warmed. Taq polymerase synthesizes new strands of DNA.

Polymerase chain reaction, or PCR, is used to amplify a specific sequence of DNA. Primers—short pieces of DNA complementary to each end of the target sequence—are combined with genomic DNA, Taq polymerase, and deoxynucleotides. Taq polymerase is a DNA polymerase isolated from the thermostable bacterium *Thermus aquaticus* that is able to withstand the high temperatures used in PCR. *Thermus aquaticus* grows in the Lower Geyser Basin of Yellowstone National Park. Reverse transcriptase PCR (RT-PCR) is similar to PCR, but cDNA is made from an RNA template before PCR begins.

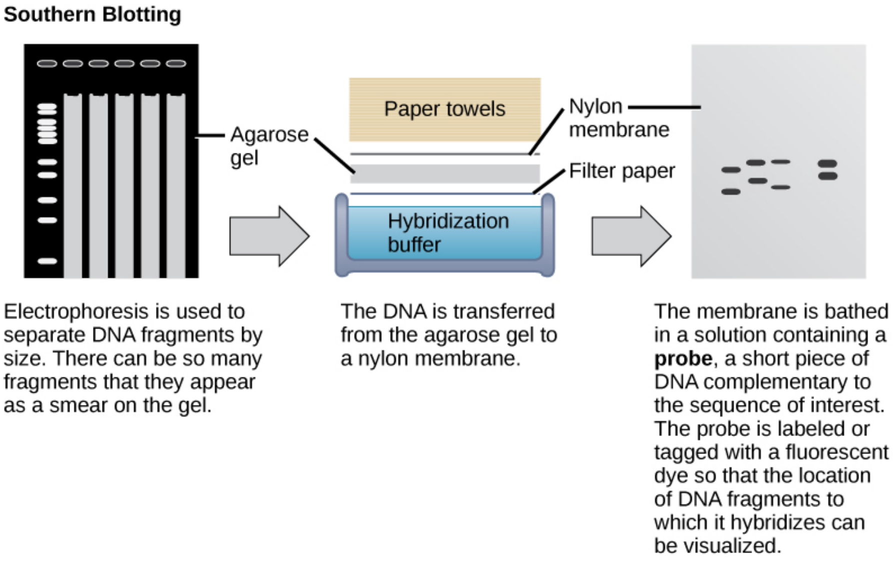
DNA fragments can also be amplified from an RNA template in a process called **reverse transcriptase PCR (RT-PCR)**. The first step is to recreate the original DNA template strand (called cDNA) by applying DNA nucleotides to the mRNA. This process is called reverse transcription. This requires the presence of an enzyme called reverse transcriptase. After the cDNA is made, regular PCR can be used to amplify it.

Note:
Link to Learning

Deepen your understanding of the polymerase chain reaction by clicking through this interactive exercise.

Hybridization, Southern Blotting, and Northern Blotting

Nucleic acid samples, such as fragmented genomic DNA and RNA extracts, can be probed for the presence of certain sequences. Short DNA fragments called **probes** are designed and labeled with radioactive or fluorescent dyes to aid detection. Gel electrophoresis separates the nucleic acid fragments according to their size. The fragments in the gel are then transferred onto a nylon membrane in a procedure called **blotting** ([link]). The nucleic acid fragments that are bound to the surface of the membrane can then be probed with specific radioactively or fluorescently labeled probe sequences. When DNA is transferred to a nylon membrane, the technique is called **Southern blotting**, and when RNA is transferred to a nylon membrane, it is called **northern blotting**. Southern blots are used to detect the presence of certain DNA sequences in a given genome, and northern blots are used to detect gene expression.


Southern Blotting
Agarose gel
Paper towels
Nylon membrane
Filter paper
Hybridization buffer
Electrophoresis is used to separate DNA fragments by size. There can be so many fragments that they appear as a smear on the gel.
The DNA is transferred from the agarose gel to a nylon membrane.
The membrane is bathed in a solution containing a **probe**, a short piece of DNA complementary to the sequence of interest. The probe is labeled or tagged with a fluorescent dye so that the location of DNA fragments to which it hybridizes can be visualized.

Southern blotting is used to find a particular sequence in a sample of DNA. DNA fragments are separated on a gel, transferred to a nylon membrane, and incubated with a DNA probe complementary to the sequence of interest. Northern blotting is similar to Southern blotting, but RNA is run on the gel instead of DNA. In western blotting, proteins are run on a gel and detected using antibodies.

Molecular Cloning

In general, the word “cloning” means the creation of a perfect replica; however, in biology, the re-creation of a whole organism is referred to as “reproductive cloning.” Long before attempts were made to clone an entire organism, researchers learned how to reproduce desired regions or fragments of the genome, a process that is referred to as molecular cloning.

Cloning small fragments of the genome allows for the manipulation and study of specific genes (and their protein products), or noncoding regions in isolation. A plasmid (also called a vector) is a small circular DNA molecule that replicates independently of the chromosomal DNA. In cloning, the plasmid molecules can be used to provide a "folder" in which to insert a desired DNA fragment. Plasmids are usually introduced into a bacterial host for proliferation. In the bacterial context, the fragment of DNA from the human genome (or the genome of another organism that is being studied) is referred to as **foreign DNA**, or a transgene, to differentiate it from the DNA of the bacterium, which is called the **host DNA**.

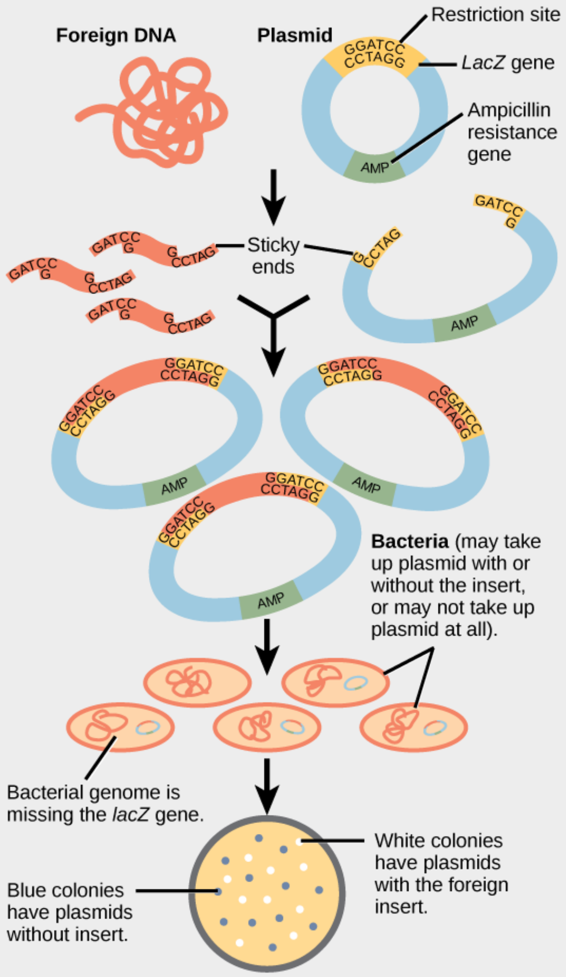
Plasmids occur naturally in bacterial populations (such as *Escherichia coli*) and have genes that can contribute favorable traits to the organism, such as **antibiotic resistance** (the ability to be unaffected by antibiotics). Plasmids have been repurposed and engineered as vectors for molecular cloning and the large-scale production of important reagents, such as insulin and human growth hormone. An important feature of plasmid vectors is the ease with which a foreign DNA fragment can be introduced via the **multiple cloning site (MCS)**. The MCS is a short DNA sequence containing multiple sites that can be cut with different commonly available restriction enzymes. **Restriction endonucleases** (enzymes) recognize specific DNA sequences and cut them in a predictable manner; they are naturally produced by bacteria as a defense mechanism against foreign DNA. Many restriction endonucleases make staggered cuts in the two strands of DNA, such that the cut ends have a 2- or 4-base single-stranded overhang. Because these overhangs are capable of annealing with complementary overhangs, these are called “sticky ends.” Addition of an enzyme called DNA ligase permanently joins the DNA fragments via phosphodiester bonds. In this way, any DNA fragment generated by restriction endonuclease cleavage can be spliced between the two ends of a plasmid DNA that has been cut with the same restriction endonuclease ([link]).

Recombinant DNA Molecules

Plasmids with foreign DNA inserted into them are called **recombinant DNA** molecules because they are created artificially and do not occur in nature. They are also called chimeric molecules because the origin of different parts of the molecules can be traced back to different species of biological organisms or even to chemical synthesis. Proteins that are expressed from recombinant DNA molecules are called **recombinant proteins**. Not all recombinant plasmids are capable of expressing genes. The recombinant DNA may need to be moved into a different vector (or host) that is better designed for gene expression. Plasmids may also be engineered to express proteins only when stimulated by certain environmental factors, so that scientists can control the expression of the recombinant proteins.

Note:
Art Connection

Molecular Cloning


Foreign DNA
Plasmid
GGATCC
CCTAGG
Restriction site
LacZ gene
Ampicillin resistance gene
AMP
GATCC
G
G
CCTAG
Sticky ends
Bacteria (may take up plasmid with or without the insert, or may not take up plasmid at all).
Bacterial genome is missing the *lacZ* gene.
White colonies have plasmids with the foreign insert.
Blue colonies have plasmids without insert.

The foreign DNA and plasmid are cut with the same **restriction enzyme**, which recognizes a particular sequence of DNA called a *restriction site*. The restriction site occurs only once in the plasmid, and is located within the *lacZ* gene, a gene necessary for metabolizing lactose.

The restriction enzyme creates sticky ends that allow the foreign DNA and cloning vector to anneal. An enzyme called ligase glues the annealed fragments together.

The ligated cloning vector is transformed into a bacterial host strain that is ampicillin sensitive and is missing the *lacZ* gene from its genome.

Bacteria are grown on media containing ampicillin and X-gal, a chemical that is metabolized by the same pathway as lactose. The ampicillin kills bacteria without plasmid. Plasmids lacking the foreign insert have an intact *lacZ* gene and are able to metabolize X-gal, releasing a dye that turns the colony blue. Plasmids with an insert have a disrupted *lacZ* gene and produce white colonies.

This diagram shows the steps involved in molecular cloning.

You are working in a molecular biology lab and, unbeknownst to you, your lab partner left the foreign genomic DNA that you are planning to clone on the lab bench overnight instead of storing it in the freezer. As a result, it was degraded by nucleases, but still used in the experiment. The plasmid, on the other hand, is fine. What results would you expect from your molecular cloning experiment?

a. There will be no colonies on the bacterial plate.
b. There will be blue colonies only.
c. There will be blue and white colonies.
d. The will be white colonies only.

Note:
Link to Learning

View an animation of recombination in cloning from the DNA Learning Center.

Gene Editing

Gene Editing is the newer, more specific, form of removing or adding genes to a DNA molecule. It is based on the Class V restriction enzymes known as **CRISPR/Cas9**. The name is an acronym derived from the description given to unusual clusters of DNA sequences first identified in bacteria in the 1970s: Clustered Regularly Interspaced Palindromic Repeats. The function of these clusters was unknown, but eventually it became clear that they form the basis of a bacterial immune process used to deactivate viruses that infect the bacterial cell. CRISPR generates an RNA sequence that forms the complement of a specific DNA sequence in the invading virus. When the RNA attaches to its complementary DNA sequence, an associated molecule cuts the DNA at the specific point of attachment. Cas9 (CRISPR-associated nuclease) is the enzyme that works with CRISPR to actually cut the DNA. When the CRISPR probe attaches to its complement on the DNA, the Cas9 enzyme is activated and cuts the DNA at the point of attachment. The CRISPR probe guides the Cas9 knife so it only cuts at sequences specific to the viral DNA and not the bacteria's own DNA.

Once the function of CRISPR/Cas9 was understood, it did not take long before synthetic CRISPR-probe sequences were developed. The CRISPR/Cas9 method of deleting and inserting DNA segments is much cheaper, faster, and more specific than the earlier use of recombinant plasmid vectors. It has rapidly become the most widely used tool for modifying DNA. An RNA CRISPR probe can be synthesized for almost any sequence and then used to delete (silence) that sequence or replace it with another synthesized sequence. Recently a CRISPR variant has been developed that can edit RNA molecules.

Because CRISPR, unlike plasmids, is so specific and can be used to edit living eukaryotic cells it is viewed as a potential medical breakthrough for treating inherited genetic diseases. While gene editing holds great potential for treating, and even curing these diseases it also carries the potential for catastrophic misuse, including the creation of designer babies, long restricted solely to science fiction. Many scientists have called for strict protocols in all CRISPR research and specifically for restrictions controlling research using CRISPR to modify humans.

Cellular Cloning

Unicellular organisms, such as bacteria and yeast, and the somatic cells of multicellular organisms, naturally produce clones of themselves when they replicate asexually by binary fission, or in the case of eukaryotes by mitosis; this is known as **cellular cloning**. The result is two (daughter) cells, each with exact replicas of the genetic material found in the original cell.

Reproductive Cloning

Reproductive cloning is a method used to make a clone or an identical copy of an entire multicellular organism. Most multicellular organisms undergo reproduction by sexual means, which involves genetic hybridization of two individuals (parents), making it impossible for generation of an identical copy or a clone of either parent. Recent advances in biotechnology have made it possible to artificially induce asexual reproduction of mammals in the laboratory.

Parthenogenesis, or “virgin birth,” occurs when an embryo grows and develops without the fertilization of the egg occurring; this is a form of asexual reproduction. An example of parthenogenesis occurs in species in which the female lays an egg and if the egg is fertilized, it is a diploid egg and the individual develops into a female; if the egg is not fertilized, it remains a haploid egg and develops into a male. The unfertilized egg is called a parthenogenic, or virgin, egg. Some insects and reptiles lay parthenogenic eggs that can develop into adults.

Sexual reproduction requires two cells; when the haploid egg and sperm cells fuse, a diploid zygote results. The zygote nucleus contains the genetic information to produce a new individual. However, early embryonic development requires the cytoplasmic material contained in the egg cell. This idea forms the basis for reproductive cloning. Therefore, if the haploid nucleus of an egg cell is replaced with a diploid nucleus from the cell of any individual of the same species (called a donor), it will become a zygote that is genetically identical to the donor. Somatic cell nuclear transfer is the technique of transferring a diploid nucleus into an enucleated egg. It can be used for either therapeutic cloning or reproductive cloning.

The first cloned animal was Dolly, a sheep who was born in 1996. The success rate of reproductive cloning at the time was very low. Dolly lived for seven years and died of respiratory complications ([link]). There is speculation that because the cell DNA belongs to an older individual, the age of the DNA may affect the life expectancy of a cloned individual. Since Dolly, several animals such as horses, bulls, and goats have been successfully cloned, although these individuals often exhibit facial, limb, and cardiac abnormalities.

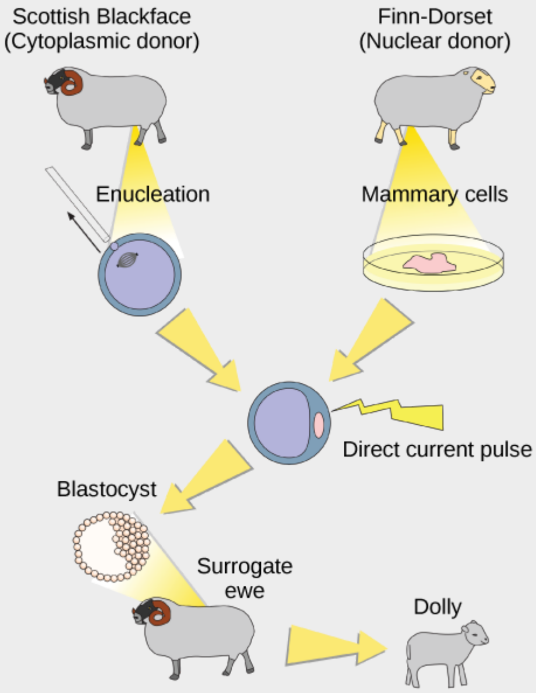
Therapeutic Cloning

There have been attempts at producing cloned human embryos as sources of embryonic stem cells, sometimes referred to as cloning for therapeutic purposes. Therapeutic cloning produces stem cells to attempt to remedy detrimental diseases or defects (unlike reproductive cloning, which aims to reproduce an organism). Still, therapeutic cloning efforts have met with resistance because of bioethical considerations.

Therapeutic cloning is not restricted to producing embryos for stem cells. Mitochondria are unique organelles in that they carry their own DNA and reproduce themselves in the cell's cytoplasm via binary fission. Thus the genetic material is already isolated and compactly packaged. This means genetic disorders involving mitochondria can be fixed by replacing the dysfunctional mitochondria with healthy mitochondria. This can be done in two ways. Both are based on the techniques used in cloning Dolly the sheep described above. The nucleus is removed from the egg of a female with dysfunctional mitochondria and implanted in an egg with healthy mitochondria that has had its own nucleus removed. The basic difference is whether the nucleus is transferred before *(Maternal Spindle Transfer)* or after *(Pronuclear Transfer)* the egg has been fertilized. In either case, cloning using the nuclei of gametes or zygotes seems to overcome the abnormalities seen in cloning mature somatic cells. In 2015 the UK became the first country to legalize this type of human cloning, resulting in what is sometimes referred to as a *three parent baby*.

Note:

Art Connection



Dolly the sheep was the first mammal to be cloned. To create Dolly, the nucleus was removed from a donor egg cell. The nucleus from a second sheep was then introduced into the cell, which was allowed to divide to the blastocyst stage before being implanted in a surrogate mother. (credit: modification of work by "Squidonius"/Wikimedia Commons)

Do you think Dolly was a Finn-Dorset or a Scottish Blackface sheep?

Genetic Engineering

Genetic engineering is the alteration of an organism’s genotype using recombinant DNA technology to modify an organism’s DNA to achieve desirable traits. Before 2012, the addition of foreign DNA in the form of recombinant DNA vectors generated by molecular cloning was the most common method of genetic engineering. However, future genetic engineering is moving rapidly toward gene editing with CRISPR/Cas9 and related technologies. Regardless of the engineering technology used, the organism that has its DNA altered is called a **genetically modified organism** (GMO). If the foreign DNA that is introduced comes from a different species, the host organism is called **transgenic**. Bacteria, plants, and animals have been genetically modified since the early 1970s for academic, medical, agricultural, and industrial purposes. In the US today testing for transgenic markers shows that GMOs are found in almost all processed foods.

Note:

Are GMOs Bad?

GMOs are viewed by some as bad, dangerous, or unnatural. However, like any technology it is the use that determines both the risks and the value of a specific transgenic organism. Virtually every food we eat has been genetically modified by tens of thousands of years of human selection and cross breeding. And transgenic gene flow is not unnatural. It has been occurring in nature for hundreds of millions of years through multiple processes collectively referred to as **Horizontal Gene Transfer**, also called *lateral gene transfer*. Bacterial plasmids, discussed above, are one of the naturally occurring vectors in bacteria, while viruses appear to be an important transgenic vector in everything from bacteria to humans. In fact, horizontal gene transfer appears to be an important evolutionary factor from the earliest days of life.

The genes for proteins that allow cells to stick together and fuse appear to be transgenic genes that were introduced into early single cells through viral infection. The implications are enormous as cell adhesion and fusion are critical for everything from multicelluarity and organ development to gamete fertilization. Approximately eight percent of the human genome consists of transgenic fragments that can be traced back to viral origins. While the functions for many of these fragments have not been identified, some have been shown to become active at critical periods. For example, in 2015 researchers showed that certain transgenic viral genes become active during the earliest stages of embryonic development. Some appear to help guide the early embryonic development while others appear to protect these young embryonic cells from infection by other viruses.

While some GMOs have been developed for what appear to be primarily corporate profit motives, others have been developed for true humanitarian reasons. One of these humanitarian GMOs is golden rice. Vitamin A deficiency is common in many areas of the world where rice is the staple carbohydrate. Each year in these areas, vitamin A deficiency is responsible for the death of over half a million children under the age of five. The genes for the beta carotene biosynthetic pathway (the precursor for vitamin A) were added to a strain of rice in 2000 as an inexpensive and efficient way to reduce this deficiency. The beta carotene it produces gives the kernels a golden color, hence the name *golden rice*. In 2005 an improved strain that produced 23 times more beta carotene was released. However, even with this clear humanitarian focus there still are protestors who object to the use of golden rice. Three different reasons seem to predominate. The first is an objection to GMOs in general that invokes the "slippery slope" argument. They argue that even if this is a "good" GMO it opens the door to others that might not be so good. The second is that golden rice does not deliver enough beta carotene to solve the problem, although the 2005 strain has been tested for vitamin A bio-availability in children and seems to overcome this objection. The third is that replacing the native varieties with a GMO strain will reduce the genetic variability and lead to potential problems with rice diseases and pests. This objection is being addressed by projects in several countries where the golden rice is being crossbred with local cultivars to increase its underlying biodiversity.

There is no doubt that the value and risks of some GMO applications are questionable. However, as the world faces increased pressure from population growth, pollution, and climate change GMOs may well be our only hope of surviving as a species. As humans that share a multinational world we are not likely going to stop GMO development, nor should we be looking at the problem with such a broad brush. What we can do is influence the direction and safety of GMOs by creating reasonable policies and developing testing regulations that focus on the value and risks of specific kinds of genetically modified organisms rather then GMOs in general.

Identifying Genes and Their Functions

Although classical methods of studying the function of genes began with a given phenotype and determined the genetic basis of that phenotype, modern techniques allow researchers to start at the DNA sequence level and ask: "What does this gene or DNA element do?" This technique, called reverse genetics, has resulted in reversing the classic genetic methodology. This method would be similar to damaging a body part to determine its function. An insect that loses a wing cannot fly, which means that the function of the wing is flight. The classical genetic method would compare insects that cannot fly with insects that can fly, and observe that the non-flying insects have lost wings. Similarly, mutating or deleting genes provides researchers with clues about gene function. The methods used to disable gene function are collectively called **gene silencing**.

In the past, gene silencing used recombinant DNA vectors to alter the expression of a particular gene, either by introducing mutations in a gene, or by eliminating the expression of a certain gene by deleting a part or all of the gene sequence from the genome of an organism. However, today most researchers have begun using CRISPR/Cas9 to silence or insert genes due to its malleability, specificity, efficiency, and cost. This does not mean there are not challenges when using CRISPR. CRISPR and Cas9 are separate molecules, so in order to work, both molecules have to be inserted into the target cell. For research purposes, a strain of experimental mice has been developed with the Cas9 gene inserted into their genome so only the CRISPER probe needs to be added, making them *edit ready*.

Gene Drive

The ability to change the genetic makeup of whole populations can take long periods of time. This is primarily because each individual in the population carries two copies of each gene. If one chromosome carries a mutated or genetically altered copy of a gene, and the other does not, the altered trait will only be passed to the next generation 50% of the time. If only a few individuals carry the trait, it may quickly be eliminated by chance before it takes hold. Even if a genetically altered individual is homozygous, once it mates with a normal individual, all of the offspring are now heterozygous and subject to the 50% problem.

However, if the altered gene is found on both chromosomes, it will be passed on to 100% of the offspring and will propagate more rapidly through the population. For hundreds of millions of years some genes have been able to cut and paste a copy of themselves onto a homologous chromosome that does not already have a copy, thereby making a heterozygous individual homozygous for that gene. This has allowed these genes to rapidly increase and persist within the population, because now 100% of the offspring in each generation will carry the altered gene, potentially forever. The process is referred to as a **gene drive** because it drives the altered gene though the population more rapidly than traditional Mendelian processes, and can also overcome all but the most potent selection pressures.

Merging CRISPR-Cas9 with gene-drive processes has opened new possibilities that carry both promise and great risk. For example, it now appears possible to genetically alter mosquitoes so the protozoan that causes malaria cannot reproduce in them. Releasing these altered mosquitoes into the wild might reduce the incidence of malaria temporarily. However, over a few generations the altered genes would likely become increasingly diluted in the population until they effectively disappeared. But if the altered genes in these mosquitoes were also altered with gene drive mechanisms and the modified mosquitoes were then released into the wild, the resistance would likely spread rapidly and might eliminate malaria altogether within a few years.

The problem is, no one knows the side effects of introducing gene-drive changes into the general environment. This is particularly concerning because the gene drive seems to insure that the gene is perpetuated in the experimental populations. It is difficult to test the actual efficacy of a gene drive experiment, because once let loose a gene drive alteration may be hard to control or stop if and when side effects emerge. There currently are proposed hypotheses on how to make the gene drive self-limiting so it disappears after a set number of generations. Additionally, recent research on fruit flies indicates that in wild populations resistance to CRIPR/cas9 gene-drive alterations may develop rapidly through naturally occurring processes. Nevertheless, many scientists have called for strong protocols on laboratory research and a ban on applying gene drive technologies in wild environments until mechanisms that control the gene drive can be proven.

Biotechnology in Medicine and Agriculture

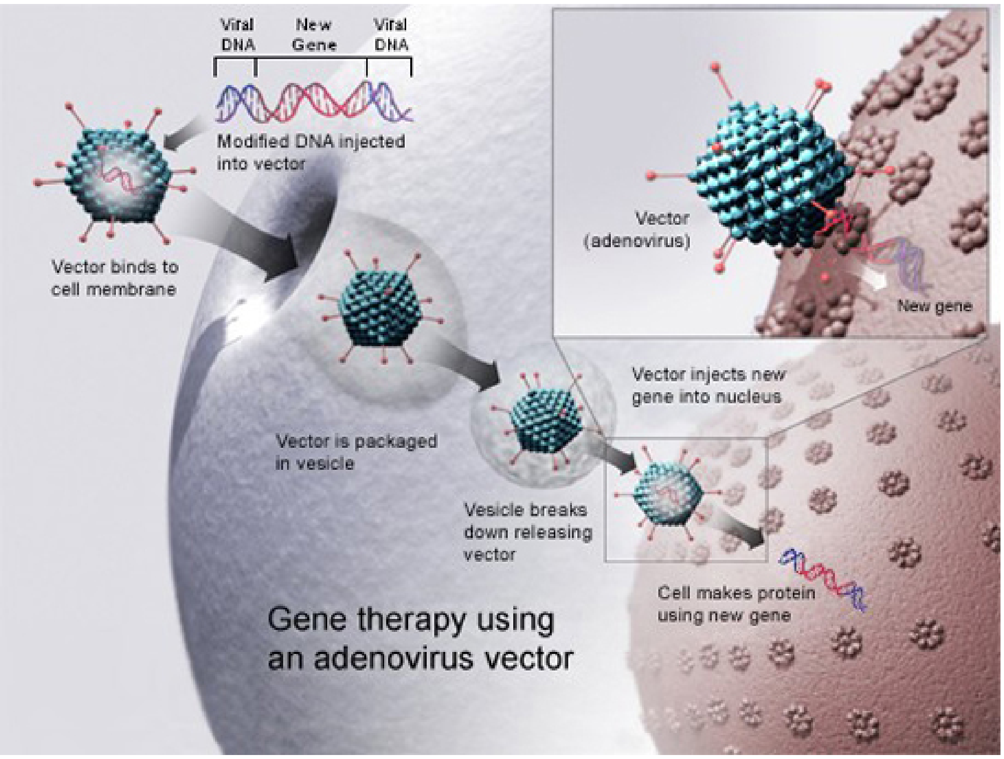
It is easy to see how biotechnology can be used for medicinal purposes. Knowledge of the genetic makeup of our species, the genetic basis of heritable diseases, and the invention of technology to manipulate and fix mutant genes provides methods to treat the disease. Biotechnology in agriculture can enhance resistance to disease, pest, and environmental stress, and improve both crop yield and quality.

Genetic Diagnosis and Gene Therapy

The process of testing for suspected genetic defects before administering treatment is called **genetic diagnosis** by **genetic testing**. Depending on the inheritance patterns of a disease-causing gene, family members are advised to undergo genetic testing. For example, women diagnosed with breast cancer are usually advised to have a biopsy so that the medical team can determine the genetic basis of cancer development. Treatment plans are based on the findings of genetic tests that determine the type of cancer. If the cancer is caused by inherited gene mutations, other female relatives are also advised to undergo genetic testing and periodic screening for breast cancer. Genetic testing is also offered for fetuses (or embryos with in vitro fertilization) to determine the presence or absence of disease-causing genes in families with specific debilitating diseases.

However, genetic diagnosis is no longer constrained to determining genetic predispositions for certain diseases. For example, a number of specific mutations can cause a cell to become cancerous. The problem is that the cancer cells in any individual often have a combination of mutations that are different even from other individuals with the same type of cancer. This is why the effectiveness of previous treatments have been so erratic. As DNA sequencing techniques become cheaper and more available, identifying the specific mutations in the cancer cells of individual cancer patients is rapidly becoming the front line for treatment. Knowing the specific mutations in this patient's cancer cells, how many and which types, allows for individualized treatments that circumvent the resistance created by each mutation.

Gene therapy is a genetic engineering technique used to cure disease. In its simplest form, it involves the introduction of a good gene at a random location in the genome to aid the cure of a disease that is caused by a mutated gene. The good gene is usually introduced into diseased cells as part of a vector transmitted by a virus that can infect the host cell and deliver the foreign DNA ([link]). More advanced forms of gene therapy try to correct the mutation at the original site in the genome, such as is the case with treatment of severe combined immunodeficiency (SCID). The CRISPR/Cas9 technology holds great promise for correcting the mutation at the original site. The biggest challenge now for gene therapy past the embryonic stage appears to be developing the mechanisms to deliver both the CRISPR probe and the Cas9 enzyme to the specific cells requiring modification. A big concern is that if Cas9 is applied systemically it may be activated by other cells in the body and cause severe side effects or initiate cancerous growth.


Viral DNA
New Gene
Viral DNA
Modified DNA injected into vector
Vector binds to cell membrane
Vector (adenovirus)
New gene
Vector injects new gene into nucleus
Vector is packaged in vesicle
Vesicle breaks down releasing vector
Cell makes protein using new gene
Gene therapy using an adenovirus vector

Gene therapy using an adenovirus vector can be used to cure certain genetic diseases in which a person has a defective gene. (credit: NIH)

Production of Vaccines, Antibiotics, and Hormones

Traditional vaccination strategies use weakened or inactive forms of microorganisms to mount the initial immune response. Modern techniques use the genes of microorganisms cloned into vectors to mass produce the desired antigen. The antigen is then introduced into the body to stimulate the primary immune response and trigger immune memory. Genes cloned from the influenza virus have been used to combat the constantly changing strains of this virus.

Antibiotics are a biotechnological product. They are naturally produced by microorganisms, such as fungi, to attain an advantage over bacterial populations. Antibiotics are produced on a large scale by cultivating and manipulating fungal cells.

Recombinant DNA technology was used to produce large-scale quantities of human insulin in *E. coli* as early as 1978. Previously, it was only possible to treat diabetes with pig insulin, which caused allergic reactions in humans because of differences in the gene product. In addition, human growth hormone (HGH) is used to treat growth disorders in children. The HGH gene was cloned from a cDNA library and inserted into *E. coli* cells by cloning it into a bacterial vector.

Transgenic Animals

Although several recombinant proteins used in medicine are successfully produced in bacteria, some proteins require a eukaryotic animal host for proper processing. For this reason, the desired genes are cloned and expressed in animals, such as sheep, goats, chickens, and mice. Animals that have been modified to express recombinant DNA are called transgenic animals. Several human proteins are expressed in the milk of transgenic sheep and goats, and some are expressed in the eggs of chickens. Mice have been used extensively for expressing and studying the effects of recombinant genes and mutations. This level of transgenic modification will likely accelerate in the future with the addition of the CRISPR/Cas9 tool set.

Transgenic Plants

Manipulating the DNA of plants (i.e., creating GMOs) has helped to create desirable traits, such as disease resistance, herbicide and pesticide resistance, better nutritional value, and better shelf-life ([link]). Plants are the most important source of food for the human population. Farmers developed ways to select for plant varieties with desirable traits long before modern-day biotechnology practices were established.

Corn, a major agricultural crop used to create products for a variety of industries, is often modified through plant biotechnology. (credit: Keith Weller, USDA)

Plants that have received recombinant DNA from other species are called transgenic plants. Because they are not natural, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because foreign genes can spread to other species in the environment, extensive testing is required to ensure ecological stability. Staples like corn, potatoes, and tomatoes were the first crop plants to be genetically engineered.

Transformation of Plants Using *Agrobacterium tumefaciens*

Gene transfer occurs naturally between species in microbial populations. Many viruses that cause human diseases, such as cancer, act by incorporating their DNA into the human genome. In plants, tumors caused by the bacterium *Agrobacterium tumefaciens* occur by transfer of DNA from the bacterium to the plant. Although the tumors do not kill the plants, they make the plants stunted and more susceptible to harsh environmental conditions. Many plants, such as walnuts, grapes, nut trees, and beets, are affected by *A. tumefaciens*. The artificial introduction of DNA into plant cells is more challenging than in animal cells because of the thick plant cell wall.

Researchers used the natural transfer of DNA from *Agrobacterium* to a plant host to introduce DNA fragments of their choice into plant hosts. In nature, the disease-causing *A. tumefaciens* have a set of plasmids, called the **Ti plasmids** (tumor-inducing plasmids), that contain genes for the production of tumors in plants. DNA from the Ti plasmid integrates into the infected plant cell's genome. Researchers manipulate the Ti plasmids to remove the tumor-causing genes and insert the desired DNA fragment for transfer into the plant genome. The Ti plasmids carry antibiotic resistance genes to aid selection and can be propagated in *E. coli* cells as well.

The Organic Insecticide *Bacillus thuringiensis*

Bacillus thuringiensis (Bt) is a bacterium that produces protein crystals during sporulation that are toxic to many insect species that affect plants. Bt toxin has to be ingested by insects for the toxin to be activated. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. Modern biotechnology has allowed plants to encode their own crystal Bt toxin that acts against insects. The crystal toxin genes have been cloned from Bt and introduced into plants. Bt toxin has been found to be safe for the environment, non-toxic to humans and other mammals, and is approved for use by organic farmers as a natural insecticide.

Flavr Savr Tomato

The first GM crop to be introduced into the market was the Flavr Savr Tomato produced in 1994. Antisense RNA technology was used to slow down the process of softening and rotting caused by fungal infections, which led to increased shelf life of the GM tomatoes. Additional genetic modification improved the flavor of this tomato. The Flavr Savr tomato did not successfully stay in the market because of problems maintaining and shipping the crop.

Section Summary

Nucleic acids can be isolated from cells for the purposes of further analysis by breaking open the cells and enzymatically destroying all other major macromolecules. Fragmented or whole chromosomes can be separated on the basis of size by gel electrophoresis. Short stretches of DNA or RNA can be amplified by PCR. Southern and northern blotting can be used to detect the presence of specific short sequences in a DNA or RNA sample. The term “cloning” may refer to cloning small DNA fragments (molecular cloning), cloning cell populations (cellular cloning), or cloning entire organisms (reproductive cloning). Genetic testing is performed to identify disease-causing genes, and gene therapy is used to cure an inheritable disease.

Transgenic organisms possess DNA from a different species, usually generated by molecular cloning techniques. Vaccines, antibiotics, and hormones are examples of products obtained by recombinant DNA technology. Transgenic plants are usually created to improve characteristics of crop plants.

Art Connections

Exercise:

Problem:

[link] You are working in a molecular biology lab and, unbeknownst to you, your lab partner left the foreign genomic DNA that you are planning to clone on the lab bench overnight instead of storing it in the freezer. As a result, it was degraded by nucleases, but still used in the experiment. The plasmid, on the other hand, is fine. What results would you expect from your molecular cloning experiment?

a. There will be no colonies on the bacterial plate.
b. There will be blue colonies only.
c. There will be blue and white colonies.
d. The will be white colonies only.

Solution:

[link] B. The experiment would result in blue colonies only.

Exercise:

Problem:

[link] Do you think Dolly was a Finn-Dorset or a Scottish Blackface sheep?

Solution:

[link] Dolly was a Finn-Dorset sheep because even though the original cell came from a Scottish blackface sheep and the surrogate mother was a Scottish blackface, the DNA came from a Finn-Dorset.

Review Questions

Exercise:

Problem:GMOs are created by ________.

a. generating genomic DNA fragments with restriction endonucleases
b. introducing recombinant DNA into an organism by any means
c. overexpressing proteins in *E. coli*.
d. all of the above

Solution:

B

Exercise:

Problem:

Gene therapy can be used to introduce foreign DNA into cells ________.

a. for molecular cloning
b. by PCR
c. of tissues to cure inheritable disease
d. all of the above

Solution:

C

Exercise:

Problem: Insulin produced by molecular cloning:

a. is of pig origin

b. is a recombinant protein
c. is made by the human pancreas
d. is recombinant DNA

Solution:

B

Exercise:

Problem: Bt toxin is considered to be ________.

a. a gene for modifying insect DNA
b. an organic insecticide produced by bacteria
c. useful for humans to fight against insects
d. a recombinant protein

Solution:

B

Exercise:

Problem: The Flavr Savr Tomato:

a. is a variety of vine-ripened tomato in the supermarket
b. was created to have better flavor and shelf-life
c. does not undergo soft rot
d. all of the above

Solution:

D

Free Response

Exercise:

Problem: Describe the process of Southern blotting.

Solution:

Southern blotting is the transfer of DNA that has been enzymatically cut into fragments and run on an agarose gel onto a nylon membrane. The DNA fragments that are on the nylon membrane can be denatured to make them single-stranded, and then probed with small DNA fragments that are radioactively or fluorescently labeled, to detect the presence of specific sequences. An example of the use of Southern blotting would be in analyzing the presence, absence, or variation of a disease gene in genomic DNA from a group of patients.

Exercise:

Problem:

A researcher wants to study cancer cells from a patient with breast cancer. Is cloning the cancer cells an option?

Solution:

Cellular cloning of the breast cancer cells will establish a cell line, which can be used for further analysis

Exercise:

Problem:

How would a scientist introduce a gene for herbicide resistance into a plant?

Solution:

By identifying an herbicide resistance gene and cloning it into a plant expression vector system, like the Ti plasmid system from *Agrobacterium tumefaciens*. The scientist would then introduce it into the plant cells by transformation, and select cells that have taken up and integrated the herbicide-resistance gene into the genome.

Exercise:

Problem:

If you had a chance to get your genome sequenced, what are some questions you might be able to have answered about yourself?

Solution:

What diseases am I prone to and what precautions should I take? Am I a carrier for any disease-causing genes that may be passed on to children?

Glossary

antibiotic resistance
ability of an organism to be unaffected by the actions of an antibiotic

biotechnology
use of biological agents for technological advancement

cellular cloning
production of identical cell populations by binary fission

clone
exact replica

foreign DNA
DNA that belongs to a different species or DNA that is artificially synthesized

gel electrophoresis
technique used to separate molecules on the basis of size using electric charge

gene targeting
method for altering the sequence of a specific gene by introducing the modified version on a vector

gene therapy
technique used to cure inheritable diseases by replacing mutant genes with good genes

genetic diagnosis
diagnosis of the potential for disease development by analyzing disease-causing genes

genetic engineering
alteration of the genetic makeup of an organism

genetic testing
process of testing for the presence of disease-causing genes

genetically modified organism (GMO)
organism whose genome has been artificially changed

host DNA
DNA that is present in the genome of the organism of interest

lysis buffer
solution used to break the cell membrane and release cell contents

molecular cloning
cloning of DNA fragments

multiple cloning site (MCS)
site that can be recognized by multiple restriction endonucleases

northern blotting
transfer of RNA from a gel to a nylon membrane

polymerase chain reaction (PCR)
technique used to amplify DNA

probe

small DNA fragment used to determine if the complementary sequence is present in a DNA sample

protease

enzyme that breaks down proteins

recombinant DNA

combination of DNA fragments generated by molecular cloning that does not exist in nature; also known as a chimeric molecule

recombinant protein

protein product of a gene derived by molecular cloning

reproductive cloning

cloning of entire organisms

restriction endonuclease

enzyme that can recognize and cleave specific DNA sequences

reverse genetics

method of determining the function of a gene by starting with the gene itself instead of starting with the gene product

reverse transcriptase PCR (RT-PCR)

PCR technique that involves converting RNA to DNA by reverse transcriptase

ribonuclease

enzyme that breaks down RNA

Southern blotting

transfer of DNA from a gel to a nylon membrane

Ti plasmid

plasmid system derived from *Agrobacterium tumifaciens* that has been used by scientists to introduce foreign DNA into plant cells

transgenic

organism that receives DNA from a different species

Energy and Metabolism

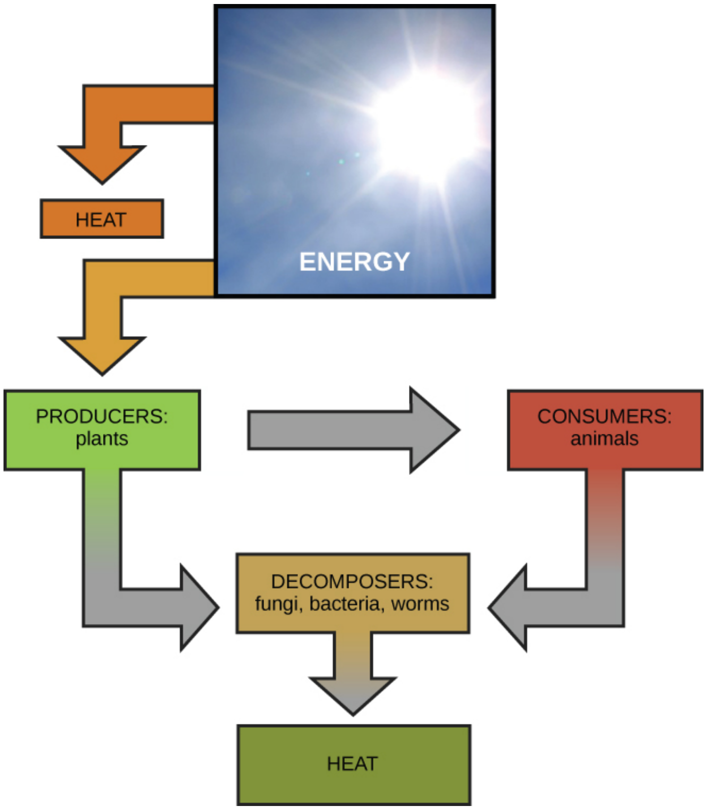
By the end of this section, you will be able to:

- Explain what metabolic pathways are
- State the first and second laws of thermodynamics
- Explain the difference between kinetic and potential energy
- Describe endergonic and exergonic reactions
- Discuss how enzymes function as molecular catalysts

As stated in the first chapter, all living organisms are dynamic kinetically stable (DKS) systems. Their structure is maintained by the constant removal and replacement of their constituent parts. This means all living organisms require a continual source of outside energy to maintain their kinetic stability. This is true at all levels of biological organization, from single cells to complex communities of interacting organisms.

The original source of outside energy for most organisms comes from sunlight, which is transformed into chemical energy by blue-green algae and plants using a process called photosynthesis. The captured chemical energy is used to maintain their own dynamic-kinetic structure. Organisms that are not capable of converting sunlight directly can gain access to the sunlight turned chemical energy by ingesting the photosynthesizing organisms and the chemical energy they contain. They retain some of this energy in their chemical structures, which in turn can be passed on to any organism that ingests them. At each level, some energy is used to maintain the organism and some is lost to the conversion processes (heat). This limits how many **trophic levels** (organisms eating organisms) can be supported before the original captured energy runs out.

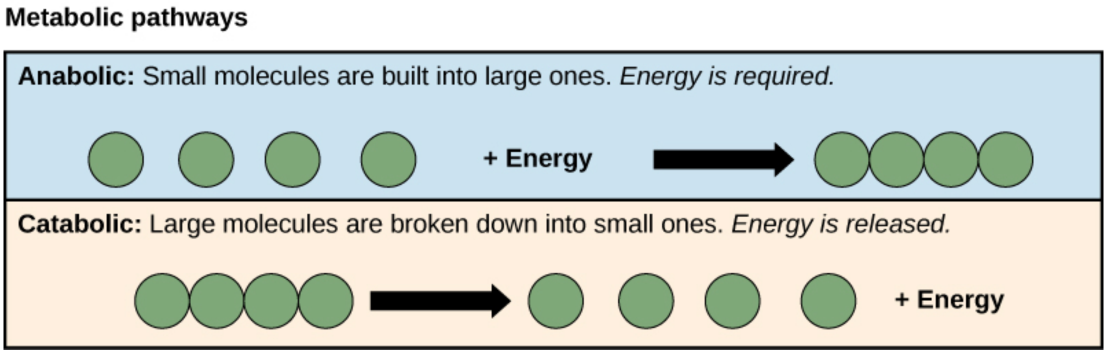
Scientists use the term **bioenergetics** to describe the concept of energy flow ([link]) through living systems, from cells to ecosystems. Cellular processes such as the building up and breaking down of complex molecules occur through stepwise chemical reactions. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. As dynamic kinetic systems, living organisms are highly ordered, and therefore require a net input of energy to maintain that order. This is why living organisms must continually consume food. They must replace the high-energy molecules that their cells use up powering the many energy-requiring chemical reactions that sustain both the highly ordered cells and the organismal infrastructure that supports them. All of these chemical reactions that take place inside cells, both those that consume energy and those that generate energy, are referred to collectively as the cell’s **metabolism**.



Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

Metabolic Pathways

A **metabolic pathway** is a series of chemical reactions that takes a starting molecule and modifies it, step-by-step much like an assembly line, through a series of metabolic intermediates, eventually yielding a final product. The processes of making and breaking down sugar molecules illustrate two examples of metabolic pathways. The first metabolic pathway synthesizes sugar from smaller molecules, and the other pathway breaks sugar down into smaller molecules. These two opposite processes—the first requiring energy and the second producing energy—are referred to as **anabolic** pathways (building polymers) and **catabolic** pathways (breaking down polymers into their monomers), respectively. Consequently, metabolism is composed of synthesis (anabolism) and degradation (catabolism) ([link]).


Metabolic pathways
Anabolic: Small molecules are built into large ones. Energy is required.
+ Energy
Catabolic: Large molecules are broken down into small ones. Energy is released.
+ Energy

Anabolic pathways are those that require energy to synthesize larger molecules. Catabolic pathways are those that generate energy by breaking down larger molecules. Both types of pathways are required for maintaining the cell’s energy balance.

Because **photosynthesis** involves synthesizing an energy-storing molecule, it requires energy input to proceed. During photosynthesis, plants use energy (originally from sunlight) to convert carbon dioxide gas (CO_2) and water (H_2O) into energy-storing sugar molecules (like glucose: $C_6H_{12}O_6$). They consume carbon dioxide and water to produce sugar molecules and the waste product oxygen (O_2). This reaction is summarized as:

Equation:

$$6CO_2 + 6H_2O \text{ --> } C_6H_{12}O_6 + 6O_2$$

Respiration is the reaction that harvests the energy stored in a sugar molecule. In cells requiring oxygen to survive respiration can be summarized as the reverse reaction to photosynthesis. In this reaction, oxygen is consumed and carbon dioxide and water are released as a waste products.

Equation:

$$C_6H_{12}O_6 + 6O_2 \text{ --> } 6H_2O + 6CO_2$$

Both of these reactions involve many steps.

During the light reactions of photosynthesis, light energy powers many individual steps that ultimately end with the creation of adenosine triphosphate (ATP) molecules, which are the primary energy currency of all cells. Just as the dollar is used as currency to buy goods, cells use molecules of ATP as energy currency to perform immediate work, such as synthesizing energy-storage molecules like glucose. Later, these energy-storing sugar molecules can be broken down and the energy stored in their chemical bonds used to synthesize new ATP molecules that can power other metabolic steps required by the organism.

It is important to know that the chemical reactions of metabolic pathways do not take place on their own. Each reaction step is facilitated, or catalyzed, by a protein called an **enzyme**. Enzymes are important for catalyzing all types of biological reactions—those that require energy as well as those that release energy.

The metabolism of sugar is but one classic example of the many cellular processes that use and produce energy. Living things consume sugars as a major energy source, because sugar molecules have a great deal of energy stored within their bonds. For the most part photosynthesizing organisms like plants produce these sugars. Metabolic processes also transform sugars (in some cases combining them with other molecules) to create and break down fatty acids, amino acids, and nucleic acids.

Energy

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter relevant to a particular case of energy transfer is called a system, and everything outside of that matter is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. In an open system, energy can be exchanged with its surroundings. The stovetop system is open because outside energy input is required to heat the pot and water and heat from the pot and water is lost to the air. A closed system cannot exchange energy with its surroundings.

Biological organisms are open systems. Energy is exchanged between them and their surroundings as they use energy from the sun to perform photosynthesis or consume energy-storing molecules and release energy to the environment by doing work and releasing heat. Like all things in the physical world, energy is subject to physical laws. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

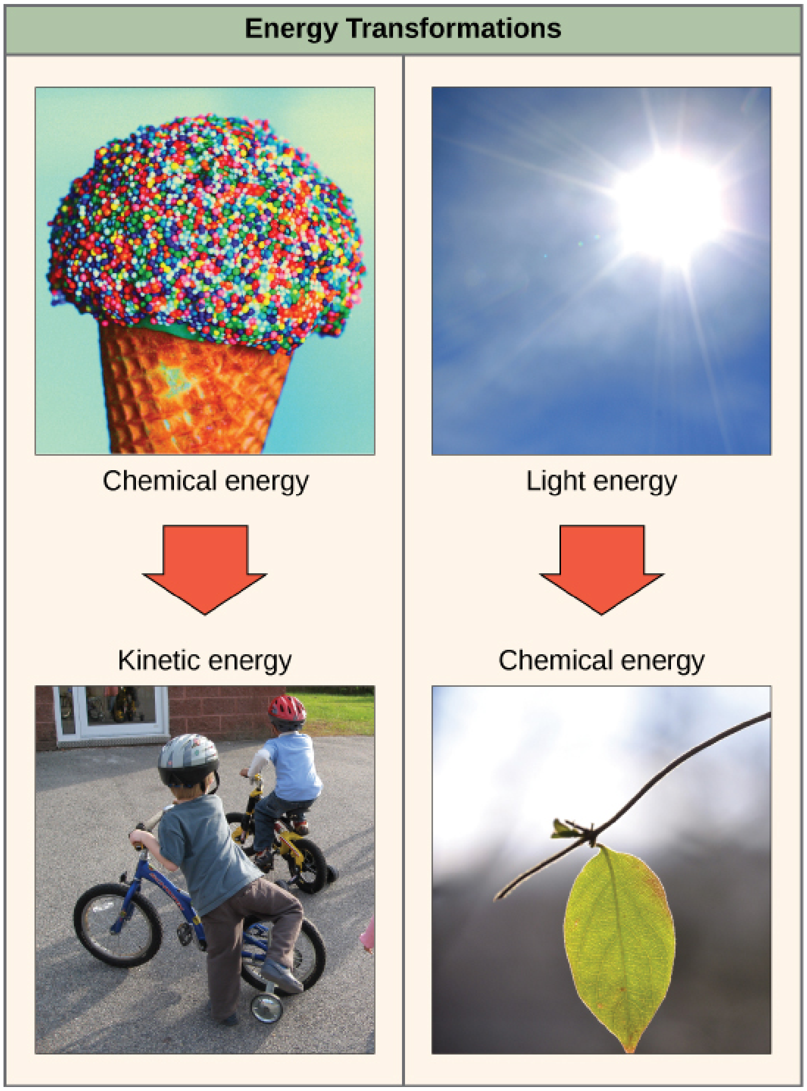
In general, energy is defined as the ability to do work, or to create some kind of change. Energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different types of energy. To appreciate the way energy flows into and out of biological systems, it is important to understand two of the physical laws that govern energy.

Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time.

Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules ([link]). Some examples of energy transformations are shown in [link].

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work to maintain their dynamic kinetic stability (life state). Living cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscle fibers to create movement.



Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit "ice cream": modification of work by D. Sharon Pruitt; credit "kids": modification of work by Max from Providence; credit "leaf": modification of work by Cory Zanker)

A living cell’s primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the second law of thermodynamics explains why these tasks are harder than they appear. All energy transfers and transformations are never completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, **heat energy** is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of randomness or disorder within a system as **entropy**. *High entropy means high disorder and low energy*. Molecules and chemical reactions have varying entropy as well. For example, entropy increases as molecules at a high concentration in one place diffuse and spread out. The second law of thermodynamics says that energy will always be lost as heat, increased entropy, in energy transfers or transformations.

Living things are highly ordered, requiring constant energy input to be maintained in a low entropy (dynamic kinetically stable) state.

Potential and Kinetic Energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. Energy associated with objects in motion is called **kinetic energy** ([link]). A speeding bullet, a walking person, and the rapid movement of molecules in the air (which produces heat) all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear, but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called **potential energy** ([link]). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.

Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. Remember that anabolic cellular pathways require energy to synthesize complex molecules from simpler ones and catabolic pathways release energy when complex molecules are broken down. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is eventually harnessed for use. This is because these bonds can release energy when broken. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called chemical energy. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

Note:
Concept in Action

Visit the site and select “Pendulum” from the “Work and Energy” menu to see the shifting kinetic and potential energy of a pendulum in motion.

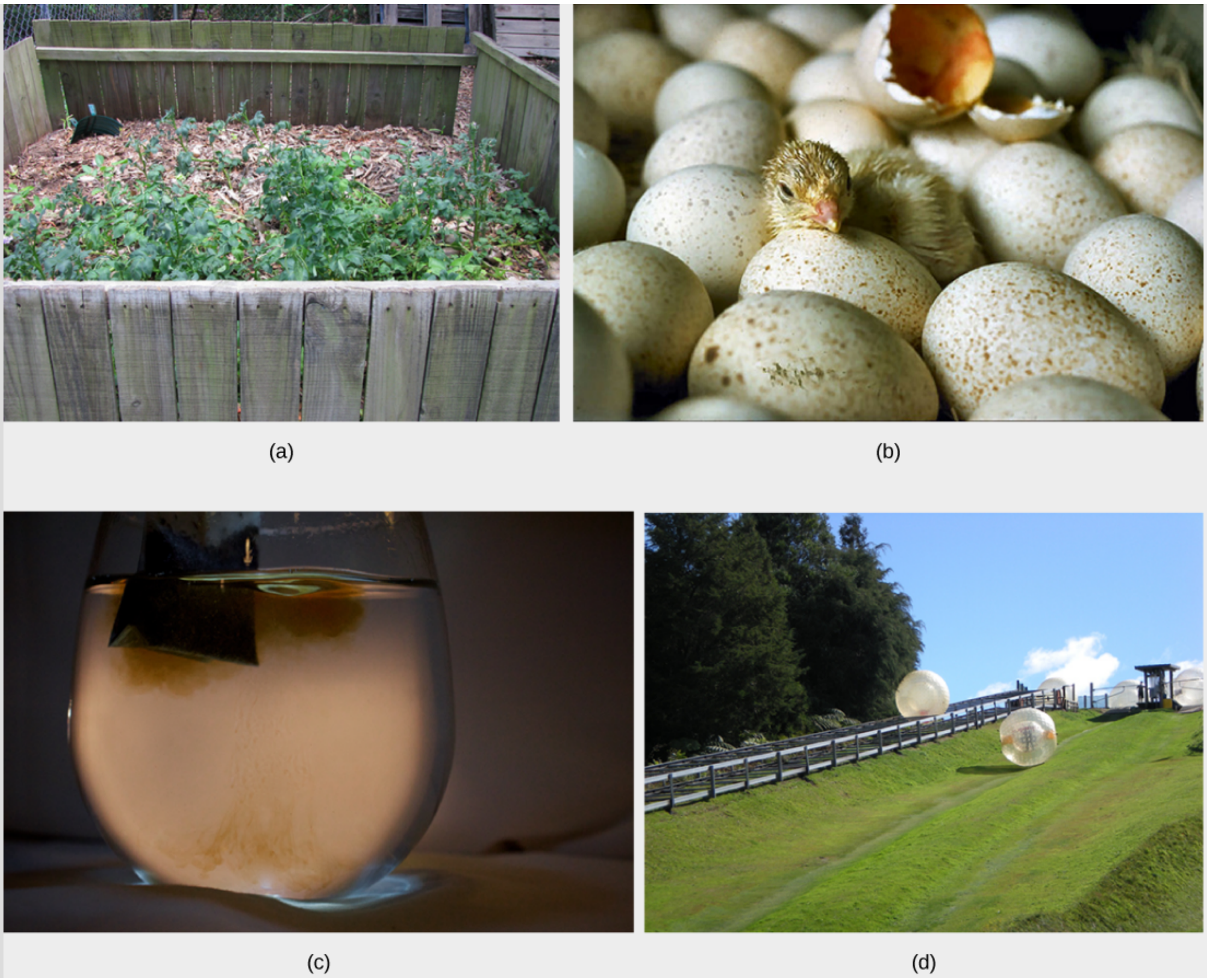
Free and Activation Energy

After learning that chemical reactions release energy when energy-storing bonds are broken, an important next question is the following: How is the energy associated with these chemical reactions quantified and expressed? How can the energy released from one reaction be compared to that of another reaction? A measurement of free energy is used to quantify these energy transfers. Recall that according to the second law of thermodynamics, all energy transfers involve the loss of some amount of energy in an unusable form such as heat. **Free energy** specifically refers to the energy associated with a chemical reaction that is available after the losses are accounted for. In other words, free energy is usable energy, or energy that is available to do work.

If energy is released during a chemical reaction, then the change in free energy, signified as ∆G (delta G) will be a negative number. A negative change in free energy also means that the products of the reaction have less free energy than the reactants, because they release some free energy during the reaction. Reactions that have a negative change in free energy and consequently release free energy are called **exergonic reactions**. Think: *ex*ergonic means energy is *ex*iting the system. These reactions are also referred to as spontaneous reactions, and their products have less stored energy than the reactants. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction occurring immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction absorbs energy rather than releases energy on balance, then the ∆G for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions** and they are non-spontaneous. An endergonic reaction will not take place on its own without the addition of free energy.

Note:
Art Connection



Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by Cory Zanker; credit d: modification of work by Harry Malsch)

Look at each of the processes shown and decide if it is endergonic or exergonic.

There is another important concept that must be considered regarding endergonic and exergonic reactions. Exergonic reactions require a small amount of energy input to get going, before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some energy input in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy**.

Note:
Concept in Action

Watch an animation of the move from free energy to transition state of the reaction.

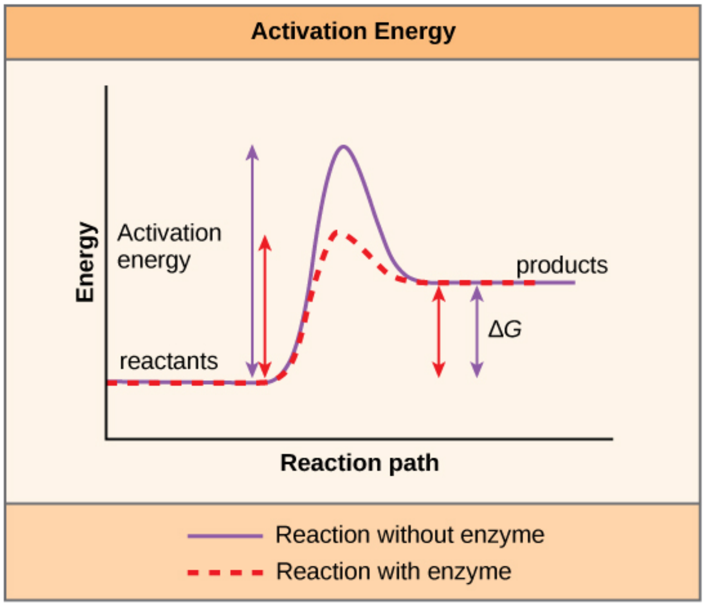
Mass Action

Many reactions are reversible, although the energy of activation often will be different depending on the direction of the reactions. The solution in which the reaction is taking place consists of reactants and product. If the process is reversible, at the molecular level reactions will continually occur in both directions as reactants form product (the forward reaction) and product breaks back down into reactants (the reverse reaction). At some point the rates of the forward and reverse reactions will reach **equilibrium**. It is important to remember that while in the solution in which the reactions are occurring (the mass action level) it will look like the amount of reactant and product is not changing, at the molecular level reactions in both directions are still happening. The concentrations of reactants and product look static at the mass level because the reaction rates in both directions cancel each other out.

The absolute concentrations of reactants and products change in a predictable manner. If more reactants are added to a solution at equilibrium they will push the reaction in a forward direction. As the new reactants are converted to product the overall concentration of product will increase in the solution. On the other hand, if more product is added to a solution at equilibrium it will push the reaction in the reverse direction. The added concentration of product will decrease as product is converted back into reactants until a new equilibrium is established. The concentration of product at the new equilibrium point will be higher than the original, but below the peak concentrations when additional product was first added. Regardless of the direction of change, the *ratio* of product to reactant concentrations will be the same at the original and new equilibrium points.

Enzymes

A substance that helps a chemical reaction to occur is called a catalyst, and the molecules that catalyze biochemical reactions are called **enzymes**. Most enzymes are proteins and perform the critical task of lowering the activation energies of chemical reactions inside the cell. Most of the reactions critical to a living cell happen too slowly at normal temperatures to be of any use to the cell. Without enzymes to speed up these reactions, life could not persist. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and -forming processes take place more easily. It is important to remember that enzymes do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the free energy of the reactants or products. They only reduce the activation energy required for the reaction to go forward ([link]). In addition, an enzyme itself is unchanged by the reaction it catalyzes. Once one reaction has been catalyzed, the enzyme is able to participate in other reactions.



Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction.

The chemical reactants to which an enzyme binds are called the enzyme’s **substrates**. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single reactant substrate is broken down into multiple products as in the catabolic processes discussed earlier. In others, two substrates may come together to create one larger molecule as in the anabolic processes discussed earlier. Two reactants might also enter a reaction and both become modified, but they leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme’s **active site**. The active site is where the “action” happens. Since enzymes are proteins, there is a unique combination of amino acid side chains within the active site. Each side chain is characterized by different properties. They can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of side chains creates a very specific chemical environment within the active site. This specific environment is suited to bind to one specific chemical substrate (or substrates).

Active sites are subject to influences of the local environment. Increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise because the molecules are moving more rapidly and thus bump into each other and interact more often. However, temperatures outside of an optimal range reduce the rate at which an enzyme catalyzes a reaction by changing the bonds maintaining the enzyme's three-dimensional shape. Hot temperatures will eventually cause enzymes to denature and can cause an irreversible change in the shape and therefore the function of the enzyme. This is the same type of irreversible change you see when an egg is cooked and the albumin protein surrounding the yolk is permanently denatured transforming from a viscous amber liquid to a white solid. Enzymes are also suited to function best within a certain pH and salt concentration range, and, as with temperature, extreme pH, and salt concentrations can affect an enzymes internal chemical bonds causing it to denature.

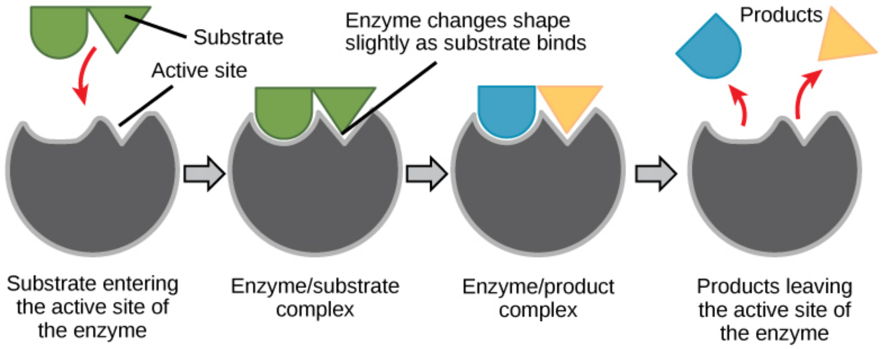
For many years, scientists thought that enzyme-substrate binding took place in a simple “lock and key” fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a model called **induced fit** ([link]). The induced-fit model expands on the lock-and-key model by describing a more dynamic binding between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme’s structure that forms an ideal binding arrangement between enzyme and substrate.

Note:
Concept in Action

View an animation of induced fit.

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of multiple possible ways. *On a basic level, enzymes promote chemical reactions that involve more than one substrate by bringing the substrates together in an optimal orientation for reaction.* Another way in which enzymes promote the reaction of their substrates is by creating an optimal environment within the active site for the reaction to occur. The chemical properties that emerge from the particular arrangement of amino acid R groups within an active site create the perfect environment for an enzyme's specific substrates to react.

The enzyme-substrate complex also can lower activation energy by compromising the bond structure so that it is easier to break. Finally, enzymes can lower activation energies by taking part in the chemical reaction itself. In these cases, it is important to remember that the enzyme will always return to its original state by the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme has catalyzed a reaction, it releases its product(s) and can catalyze a new reaction.


Substrate
Active site
Enzyme changes shape slightly as substrate binds
Products
Substrate entering the active site of the enzyme
Enzyme/substrate complex
Enzyme/product complex
Products leaving the active site of the enzyme

The induced-fit model is an adjustment to the lock-and-key model and explains how enzymes and substrates undergo dynamic modifications during the transition state to increase the affinity of the substrate for the active site.

Catalytic Rate

Catalytic rate refers to the amount of product that is produced per unit of time. Often the catalytic rate is expressed in terms of the time it takes a given concentration of reactants to reach equilibrium with the resulting product. A number of individual factors affect the rate of an enzyme mediated reaction.

The higher the concentration of enzyme, the greater the reaction rate. As long as the substrate concentration is sufficient to occupy all the enzyme active sites, the addition of more enzyme will increase the reaction rate. The rates of many reactions in the body are controlled by the amount of enzyme available.

The greater the affinity between the active site and the substrate the greater the reaction rate. Another way of saying this is the better the induced fit, the faster the reaction will occur. This is an important concept that we will encounter throughout physiology, not just in enzymatic reactions. The same type of molecule often "fits" the active site on a number of different receptors in the body. However, the affinity of each receptor for that molecule often is different. The molecule will bind more quickly with those active sites for which is has a higher affinity. This is why the dosage of therapeutic drugs can be extremely important. The higher the concentration of the drug in the body the more likely it will saturate the intended receptors and begin binding to non-targeted receptors at higher rates causing unintended side effects.

As has already been mentioned above, enzymes have optimum temperature, pH, and salt concentration ranges in which they can operate. As the conditions of any of these variables approaches the operating limit for a given enzyme the reaction rate will slow down. This generally is because the environmental change affects the three-dimensional shape of the enzyme (protein) and thus affects the affinity of its active site. An example of this can be seen in the enzymes that breakdown proteins in the digestive system. The enzyme active in the stomach only works in very low pH (acidic) environments. A separate enzyme with a different optimal range is required in the duodenum where the pH is much higher.

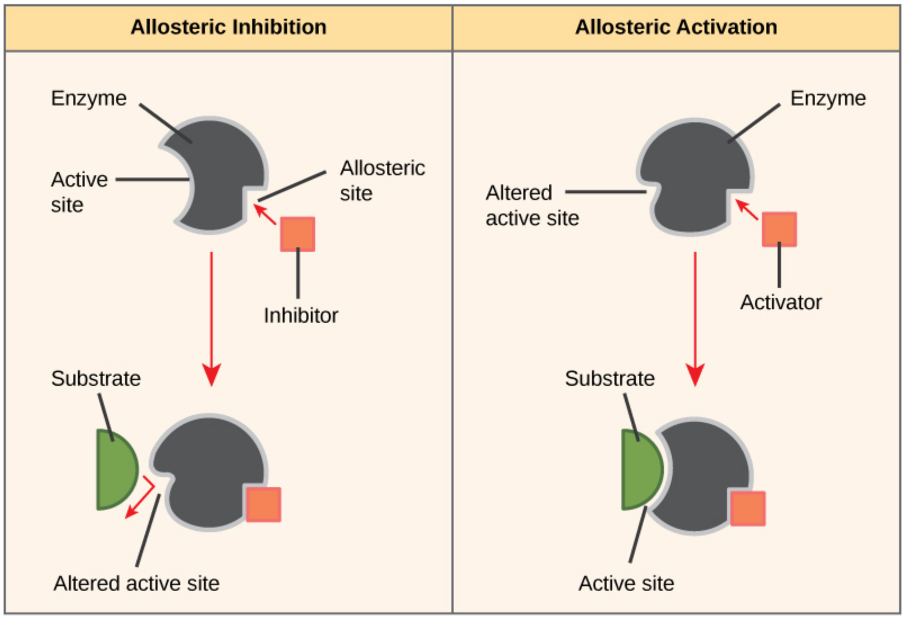
Regulation of Enzyme Activity

It would seem ideal to have a scenario in which all of an organism's enzymes existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. However, a variety of mechanisms ensures that this does not happen. Cellular needs and conditions constantly vary from cell to cell, and change within individual cells over time. The required enzymes of stomach cells differ from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive organ cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so must the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at what rates. This determination is tightly controlled in cells. In certain cellular environments, enzyme activity is partly controlled by environmental factors like pH, temperature, salt concentration, and, in some cases, cofactors or coenzymes. **Cofactors** are inorganic molecules. Most are ions like iron or magnesium; however some can be larger inorganic molecules like iron-sulfides. **Coenzymes** are organic molecules, many derived from vitamins.

Enzymes can also be regulated in ways that either promote or reduce enzyme activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms by which they do so. In some cases of enzyme inhibition, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for binding to the active site.

On the other hand, in **noncompetitive inhibition**, an inhibitor molecule binds to the enzyme in a location other than the active site, called an **allosteric site**, but still manages to block substrate binding to the active site. Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called **allosteric inhibition** ([link]). Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to a region on an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates with less efficiency. There are allosteric activators as well as inhibitors. **Allosteric activators** bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s) ([link]).



Allosteric inhibition works by indirectly inducing a conformational change to the active site such that the substrate no longer fits. In contrast, in allosteric activation, the activator molecule modifies the shape of the active site to allow a better fit of the substrate.

Note:

Careers in Action

Pharmaceutical Drug Developer

Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

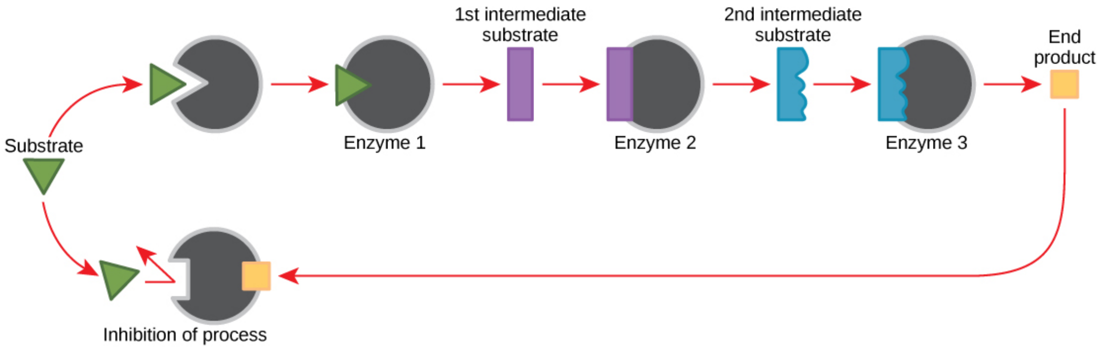
Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated are key principles behind the development of many of the pharmaceutical drugs on the market today. Biologists working in this field collaborate with other scientists to design drugs ([link]).

Consider statins for example—statins is the name given to one class of drugs that can reduce cholesterol levels. These compounds are inhibitors of the enzyme HMG-CoA reductase, which is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the level of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is used to provide relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs discovered? One of the biggest challenges in drug discovery is identifying a drug target. A drug target is a molecule that is literally the target of the drug. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the target alone is not enough; scientists also need to know how the target acts inside the cell and which reactions go awry in the case of disease. Once the target and the pathway are identified, then the actual process of drug design begins. In this stage, chemists and biologists work together to design and synthesize molecules that can block or activate a particular reaction. However, this is only the beginning: If and when a drug prototype is successful in performing its function, then it is subjected to many tests from in vitro experiments to clinical trials before it can get approval from the U.S. Food and Drug Administration to be on the market.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. The major question remains, however: What are these molecules and where do they come from? Some are cofactors and coenzymes, as you have learned. What other molecules in the cell provide enzymatic regulation such as allosteric modulation, and competitive and non-competitive inhibition? Perhaps the most relevant sources of regulatory molecules, with respect to enzymatic cellular metabolism, are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use the products of their own reactions as negative feedback to control further enzyme activity and product generation. **Feedback inhibition** refers to enzyme pathways that use a reaction product to regulate its own further production ([link]). The cell responds to an abundance of the products by slowing down production during anabolic or catabolic reactions. Such reaction products use the mechanisms described above to inhibit the enzymes that catalyzed their production .



Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream process, is an important regulatory mechanism in cells.

The production of both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that creates ATP. In this way, when ATP is in abundant supply, the cell can prevent the production of ATP. On the other hand, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP through sugar catabolism.

Section Summary

Cells perform the functions of life through various chemical reactions. A cell's metabolism refers to the combination of chemical reactions that take place within it. Catabolic reactions break down complex chemicals into simpler ones and are associated with energy release. Anabolic processes build complex molecules out of simpler ones and require energy.

In studying energy, the term system refers to the matter and environment involved in energy transfers. Entropy is a measure of the disorder of a system. The physical laws that describe the transfer of energy are the laws of thermodynamics. The first law states that the total amount of energy in the universe is constant. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy. Energy comes in different forms: kinetic, potential, and free. The change in free energy of a reaction can be negative (releases energy, exergonic) or positive (consumes energy, endergonic). All reactions require an initial input of energy to proceed, called the activation energy.

Enzymes are chemical catalysts that speed up chemical reactions by lowering their activation energy. Enzymes have an active site with a unique chemical environment that fits particular chemical reactants for that enzyme, called substrates. Enzymes and substrates are thought to bind according to an induced-fit model. Enzyme action is regulated to conserve resources and respond optimally to the environment.

Art Connections

Exercise:

Problem:

[link] Look at each of the processes shown and decide if it is endergonic or exergonic.

Solution:

[link] A compost pile decomposing is an exergonic process. A baby developing from a fertilized egg is an endergonic process. Tea dissolving into water is an exergonic process. A ball rolling downhill is an exergonic process.

Review Questions

Exercise:

Problem:

Which of the following is not an example of an energy transformation?

a. Heating up dinner in a microwave
b. Solar panels at work
c. Formation of static electricity
d. None of the above

Solution:

D

Exercise:

Problem: Which of the following is not true about enzymes?

a. They are consumed by the reactions they catalyze.
b. They are usually made of amino acids.
c. They lower the activation energy of chemical reactions.
d. Each one is specific to the particular substrate(s) to which it binds.

Solution:

A

Free Response

Exercise:

Problem:

Does physical exercise to increase muscle mass involve anabolic and/or catabolic processes? Give evidence for your answer.

Solution:

Physical exercise involves both anabolic and catabolic processes. Body cells break down sugars to provide ATP to do the work necessary for exercise, such as muscle contractions. This is catabolism. Muscle cells also must repair muscle tissue damaged by exercise by building new muscle. This is anabolism.

Exercise:

Problem:

Explain in your own terms the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods of time due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:

Problem:

With regard to enzymes, why are vitamins and minerals necessary for good health? Give examples.

Solution:

Most vitamins and minerals act as cofactors and coenzymes for enzyme action. Many enzymes require the binding of certain cofactors or coenzymes to be able to catalyze their reactions. Since enzymes catalyze many important reactions, it is critical to obtain sufficient vitamins and minerals from diet and supplements. Vitamin C (ascorbic acid) is a coenzyme necessary for the action of enzymes that build collagen.

Glossary

activation energy
the amount of initial energy necessary for reactions to occur

active site
a specific region on the enzyme where the substrate binds

allosteric inhibition
the mechanism for inhibiting enzyme action in which a regulatory molecule binds to a second site (not the active site) and initiates a conformation change in the active site, preventing binding with the substrate

anabolic
describes the pathway that requires a net energy input to synthesize complex molecules from simpler ones

bioenergetics
the concept of energy flow through living systems

catabolic
describes the pathway in which complex molecules are broken down into simpler ones, yielding energy as an additional product of the reaction

competitive inhibition
a general mechanism of enzyme activity regulation in which a molecule other than the enzyme's substrate is able to bind the active site and prevent the substrate itself from binding, thus inhibiting the overall rate of reaction for the enzyme

endergonic
describes a chemical reaction that results in products that store more chemical potential energy than the reactants

enzyme
a molecule that catalyzes a biochemical reaction

exergonic

describes a chemical reaction that results in products with less chemical potential energy than the reactants, plus the release of free energy

feedback inhibition

a mechanism of enzyme activity regulation in which the product of a reaction or the final product of a series of sequential reactions inhibits an enzyme for an earlier step in the reaction series

heat energy

the energy transferred from one system to another that is not work

kinetic energy

the type of energy associated with objects in motion

metabolism

all the chemical reactions that take place inside cells, including those that use energy and those that release energy

noncompetitive inhibition

a general mechanism of enzyme activity regulation in which a regulatory molecule binds to a site other than the active site and prevents the active site from binding the substrate; thus, the inhibitor molecule does not compete with the substrate for the active site; allosteric inhibition is a form of noncompetitive inhibition

potential energy

the type of energy that refers to the potential to do work

substrate

a molecule on which the enzyme acts

thermodynamics

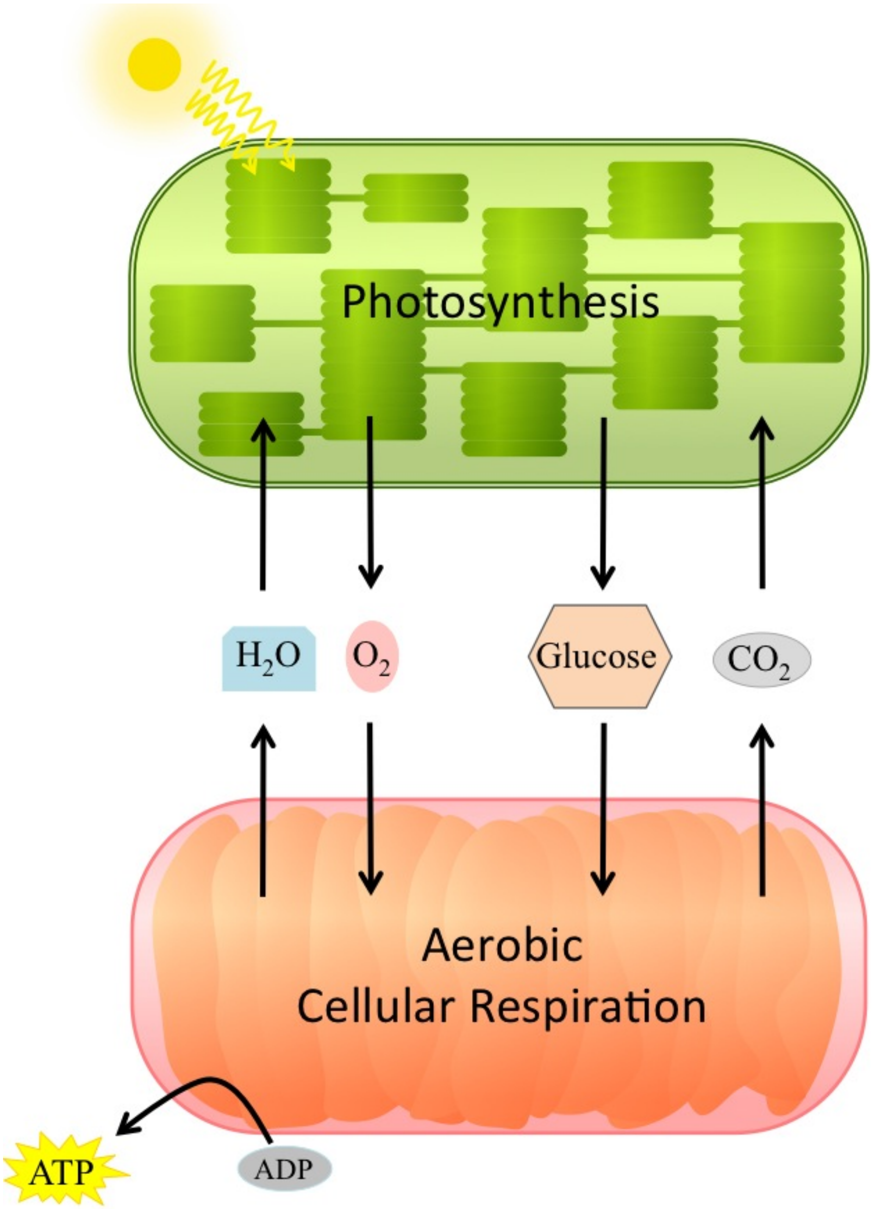
the science of the relationships between heat, energy, and work

Overview of Cellular Respiration
This is a general overview of cellular respiration.

Introduction

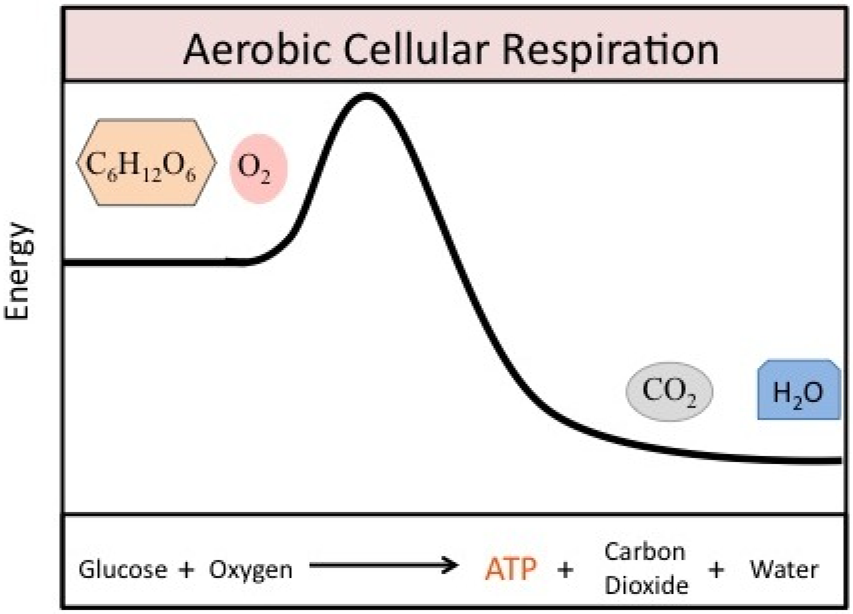
"Surely the mitochondrion that first entered another cell was not thinking about the future benefits of cooperation and integration; it was merely trying to make its own living in a tough Darwinian world." Stephen Jay Gould, in *Wonderful Life: the Burgess Shale and the Nature of History*, (1990)

All living organisms require energy, and for all organisms this energy comes from the chemical energy found in compounds that they acquire from their environment. The mitochondrion, a descendent of an aerobically-respiring bacteria, is the site of energy generation in eukaryotes. As we learned previously, the process of photosynthesis uses solar energy (sunlight) and converts this energy into chemical energy in the form of carbohydrates. In order for the chemical energy in the carbohydrates to be made available to do cellular work, the energy must be converted into a useable form known as ATP. Adenosine Triphosphate is the energy currency of the cell, and everything you do from walking down the street to reading this book requires energy in the form of ATP. Organisms need a constant supply of ATP, and the potential energy stored in food is the source of energy to meet this need. By connecting all this together, you should realize that your daily activities are fueled by the energy from the sun and that even on the cellular level nutrients cycle and energy flows ([link]).



This image illustrates the relationship between photosynthesis and cellular respiration. (Image by Eva Horne and Robert Bear)

All organisms need ATP, but not all organisms use the same pathways to generate ATP from the food that is consumed. **Aerobic cellular respiration**, the main subject of this chapter, uses oxygen (O_2) and glucose to generate ATP. Organisms (plants, animals, fungi and microbes) that live in an oxygen (O_2) rich environment use this process to generate ATP. The overall equation for aerobic cellular respiration is the reverse of photosynthesis, is an exergonic reaction, and supplies the ATP for cellular functions ([link]).


Aerobic Cellular Respiration
Energy
$C_6H_{12}O_6$
O_2
CO_2
H_2O
Glucose + Oxygen → ATP + Carbon Dioxide + Water

This image illustrates the overall equation for aerobic cellular respiration and how the amounts of free energy differs between the reactants and the products. (Image by Robert Bear)

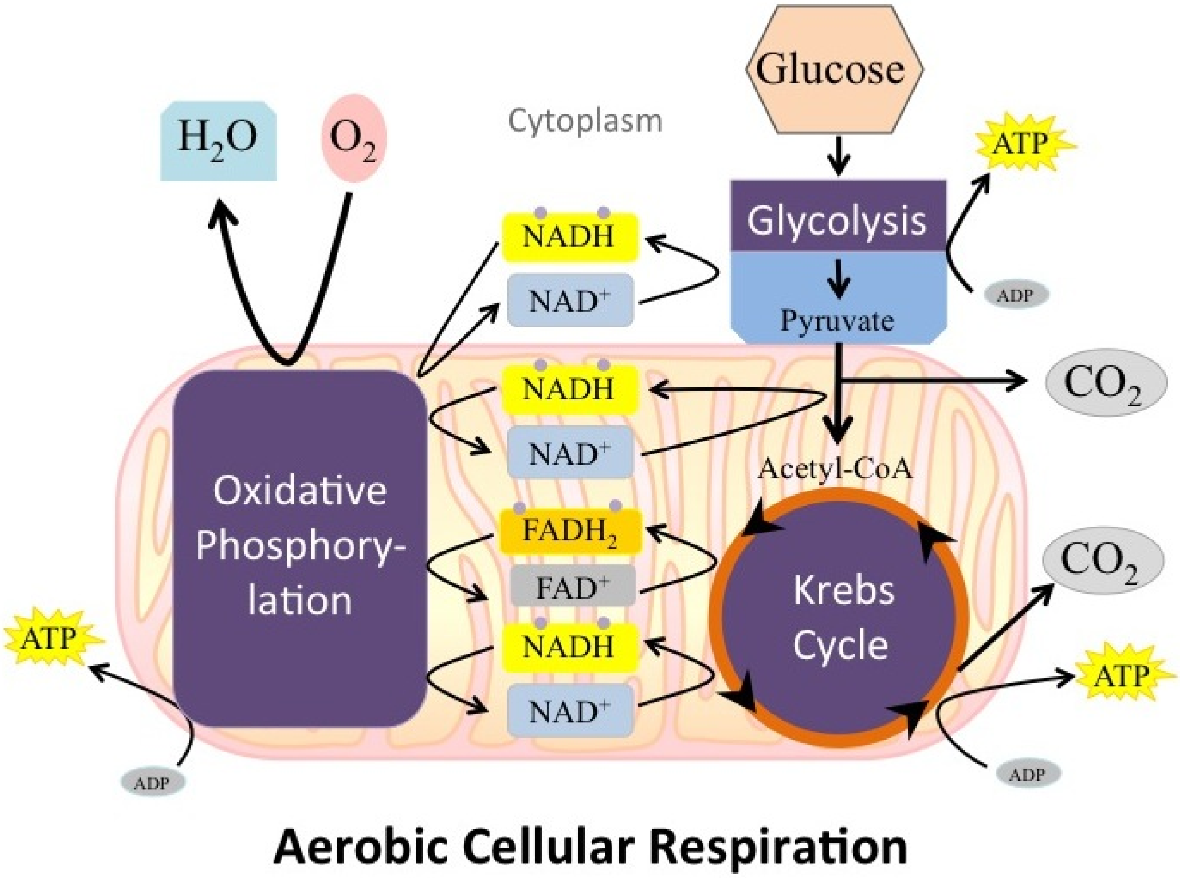
As the aerobic cellular respiration equation shows ([link]), an organism needs to acquire the O_2 from its surroundings and to get rid of the CO_2 that is produced. The acquisition of O_2 and the release of CO_2 is accomplished in a variety of ways. In single celled organisms, the movement of O_2 and CO_2 (gas exchange) is done by simple diffusion. However, in complex organisms there are specialized organs that allow for gas exchange; for example, gills in aquatic organisms and lungs in terrestrial animals.

A common misconception is that plants do not undergo cellular respiration because they make their own energy by photosynthesis. Plants do perform cellular respiration using the carbohydrates produced via photosynthesis; this occurs in tissues that are not photosynthetically active (e.g., roots), as well as in leaves and stems. Approximately half of the glucose produced by photosynthesis is consumed by the plant, mostly to generate ATP during aerobic cellular respiration. Other uses of glucose in the plant include synthesis of cell walls, starch, and other plant carbohydrates. So, plants harvest light energy via photosynthesis, making carbohydrates, and then they use the energy stored in those carbohydrates to perform various cellular functions. This is the reason why they are called **autotrophs**, or self feeders.

Some single-celled organisms use **anaerobic metabolism** to extract energy from biological molecules; this process occurs in the absence of oxygen. In this chapter, we will explore one type of anaerobic metabolism called fermentation. You may already be familiar with a one type of fermentation, lactic acid fermentation, especially if you have recently over-exerted your muscles. Anaerobic metabolism is used by many organisms to produce ATP when oxygen is not available and thus pathways which require oxygen cannot be used. The amount of ATP produced by fermentation is much less then that produced by aerobic cellular respiration, so there is a cost and benefit associated with organisms utilizing fermentation.

Summary of Aerobic Cellular Respiration

Aerobic cellular respiration ([link]) is series of linked chemical reactions that can be best understood if it is separated into four stages. These are **glycolysis**, pyruvate oxidation, the **Krebs Cycle**, and **oxidative phosphorylation**. Similar to photosynthesis, cellular respiration uses a series of oxidation-reduction reactions. During these reactions, electrons are stripped from the chemical bonds of the original glucose molecule and eventually added to oxygen, via a series of intermediate steps. This series of reactions releases small amounts of energy at each step; this energy is used to drive the formation of ATP. This section is a brief introduction to the stages of aerobic respiration with more detail to follow in the chapter.



This image illustrates aerobic cellular respiration. (Image by Eva Horne and Robert Bear)

The first stage of cellular respiration is called **Glycolysis** and occurs in the cytoplasm of the cell. During glycolysis, 1 glucose molecule (with 6 carbon atoms) is broken down into 2 pyruvate molecules (with three carbon atoms each). This is accompanied by the production of a few ATP molecules and the storage of some high-energy electrons on the electron carrier NADH. Note that no O_2 is needed for this set of reactions, which means that glycolysis can proceed in the absence of oxygen.

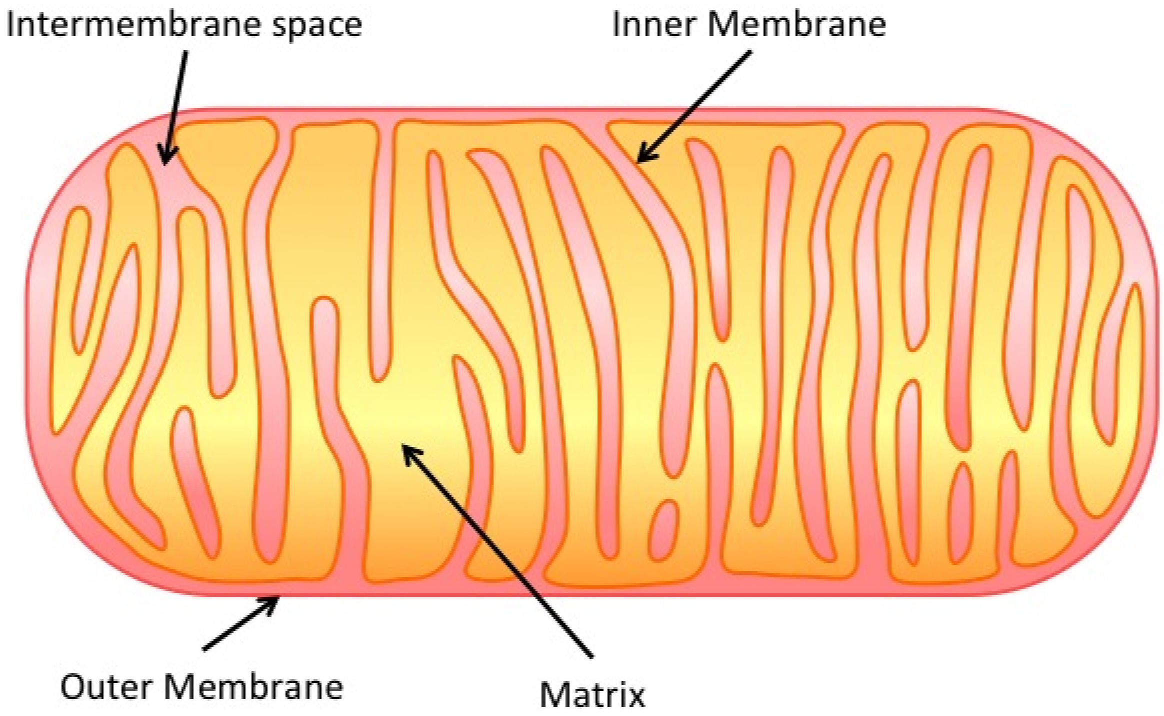
The second stage is a short series of reactions called the **oxidation of pyruvate** during which pyruvate (3 carbon atoms) is converted to acetyl-CoA (two carbon atoms), accompanied by the production of CO_2 (one carbon atom). This process occurs on the mitochondrial inner membrane, and as a result the acetyl-CoA is formed inside the mitochondria. Pyruvate is made in the cytoplasm, and this step moves the next compound in the pathway into the mitochodria. This is critical, since all subsequent steps in the pathway occur within the mitochondria. The other important event of this stage is the addition of high-energy electrons to NAD^+, generating another molecule of the electron carrier NADH.

Acetyl-CoA enters into the **Krebs cycle**, a series of mitochondrial reactions that completes the breakdown of the original glucose, thereby releasing CO_2. In this third stage of the process, energy is harvested in the form of high-energy electrons being used to generate NADH as well as another high-energy electron carrier, $FADH_2$. The reactions of the Krebs cycle also produce a small amount of ATP.

So far, a minimal amount of ATP has been produced, but a lot of energy has been stored in the electron carriers NADH and $FADH_2$. In the final stage of aerobic cellular respiration, **Oxidative Phosphorylation**, a series of enzymes known as the **electron transport chain** uses those high-energy electrons to produce a large amount of ATP. The high energy electrons harvested in the first three stages, and ferried by electron carriers (NADH and $FADH_2$) to the electron transport chain, are used to produce large amounts of ATP vir the mitochondrial membrane protein known as the **ATP synthase**. During this final stage is also when atmospheric oxygen is used as the final electron and hydrogen ion acceptor, in a reaction which produces water. The need for O_2 in this final step means that these reactions are part of aerobic cellular respiration.

Location and Structures of Aerobic Cellular Respiration

All eukaryotic cells (protists, fungi, plants and animals) have mitochondria, and mitochondria are often called the power plants of the cell because these organelles produce a large amount of ATP. As you may remember from a previous module, the mitochondrion is an organelle that is hypothesized to have originated as an endosymbiotic aerobic bacteria. Some of the evidence for this hypothesis comes from the relationship of the functional parts of the mitochondria ([link]) to the structure of a typical aerobic bacteria. There is an **outer membrane** which defines the organelle and represents the membrane which enveloped the bacteria when it was taken into the cell via endocytosis. The inner membrane represents the plasma membrane of the bacteria; the inner and outer membranes together form the **intermembrane space**. The **inner membrane** is highly folded; these folds are called **cristae**. The extensive folding increases the surface area for the numerous electron transport chain enzymes and the ATP synthases that are used to make ATP. In bacteria all of these enzymes are packed into the plasma membrane, as one would expect if the endosymbiotic hypothesis is correct. The production of ATP is driven by a concentration gradient between the outer and inner compartment; in aerobic bacteria this concentration gradient is between the inside and the outside of the cell. The innermost compartment, derived from the cytoplasm of the ancestral bacteria, is called the **matrix**, and this compartment (just like the cytoplasm of today's bacteria) contains ribosomes and DNA; It is also the location of the Krebs Cycle reactions.



This image illustrates the structures within the mitochondria. (Image by Eva Horne and Robert Bear)

Glycolysis
By the end of this section, you will be able to:

- Describe the overall result in terms of molecules produced in the breakdown of glucose by glycolysis
- Compare the output of glycolysis in terms of ATP molecules and NADH molecules produced

You have read that nearly all of the energy used by living cells comes to them in the bonds of the sugar, glucose. **Glycolysis** is the first step in the breakdown of glucose to extract energy for cellular metabolism. Nearly all living organisms carry out glycolysis as part of their metabolism. The process does not use oxygen and is therefore **anaerobic**. Glycolysis takes place in the cytoplasm of both prokaryotic and eukaryotic cells. Glucose enters heterotrophic cells in two ways. One method is through secondary active transport in which the transport takes place against the glucose concentration gradient. The other mechanism uses a group of integral proteins called GLUT proteins, also known as glucose transporter proteins. These transporters assist in the facilitated diffusion of glucose.

Glycolysis begins with the six carbon ring-shaped structure of a single glucose molecule and ends with two molecules of a three-carbon sugar called **pyruvate**. Glycolysis consists of two distinct phases. The first part of the glycolysis pathway traps the glucose molecule in the cell and uses energy to modify it so that the six-carbon sugar molecule can be split evenly into the two three-carbon molecules. The second part of glycolysis extracts energy from the molecules and stores it in the form of ATP and NADH, the reduced form of NAD.

First Half of Glycolysis (Energy-Requiring Steps)

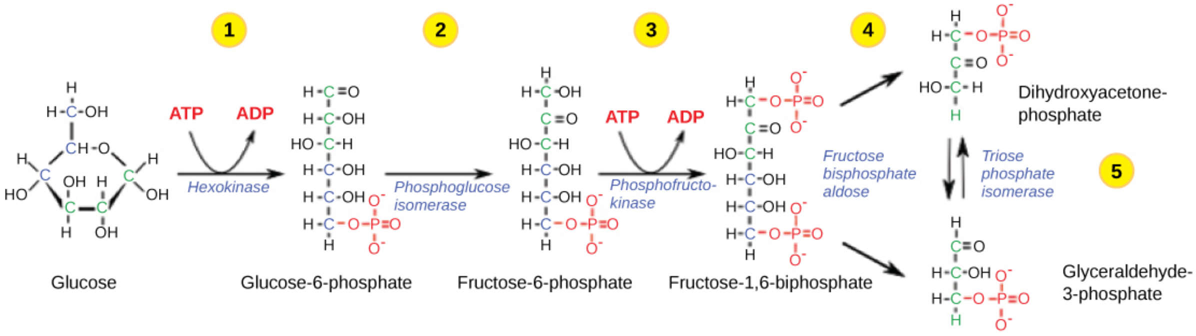
Step 1. The first step in glycolysis ([link]) is catalyzed by hexokinase, an enzyme with broad specificity that catalyzes the phosphorylation of six-carbon sugars. Hexokinase phosphorylates glucose using ATP as the source of the phosphate, producing glucose-6-phosphate, a more reactive form of glucose. This reaction prevents the phosphorylated glucose molecule from continuing to interact with the GLUT proteins, and it can no longer leave the cell because the negatively charged phosphate will not allow it to cross the hydrophobic interior of the plasma membrane.

Step 2. In the second step of glycolysis, an isomerase converts glucose-6-phosphate into one of its isomers, fructose-6-phosphate. An **isomerase** is an enzyme that catalyzes the conversion of a molecule into one of its isomers. (This change from phosphoglucose to phosphofructose allows the eventual split of the sugar into two three-carbon molecules.).

Step 3. The third step is the phosphorylation of fructose-6-phosphate, catalyzed by the enzyme phosphofructokinase. A second ATP molecule donates a high-energy phosphate to fructose-6-phosphate, producing fructose-1,6-bisphosphate. In this pathway, phosphofructokinase is a rate-limiting enzyme. It is active when the concentration of ADP is high; it is less active when ADP levels are low and the concentration of ATP is high. Thus, if there is “sufficient” ATP in the system, the pathway slows down. This is a type of end product inhibition, since ATP is the end product of glucose catabolism.

Step 4. The newly added high-energy phosphates further destabilize fructose-1,6-bisphosphate. The fourth step in glycolysis employs an enzyme, aldolase, to cleave 1,6-bisphosphate into two three-carbon isomers: dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate.

Step 5. In the fifth step, an isomerase transforms the dihydroxyacetone-phosphate into its isomer, glyceraldehyde-3-phosphate. Thus, the pathway will continue with two molecules of a single isomer. At this point in the pathway, there is a net investment of energy from two ATP molecules in the breakdown of one glucose molecule.

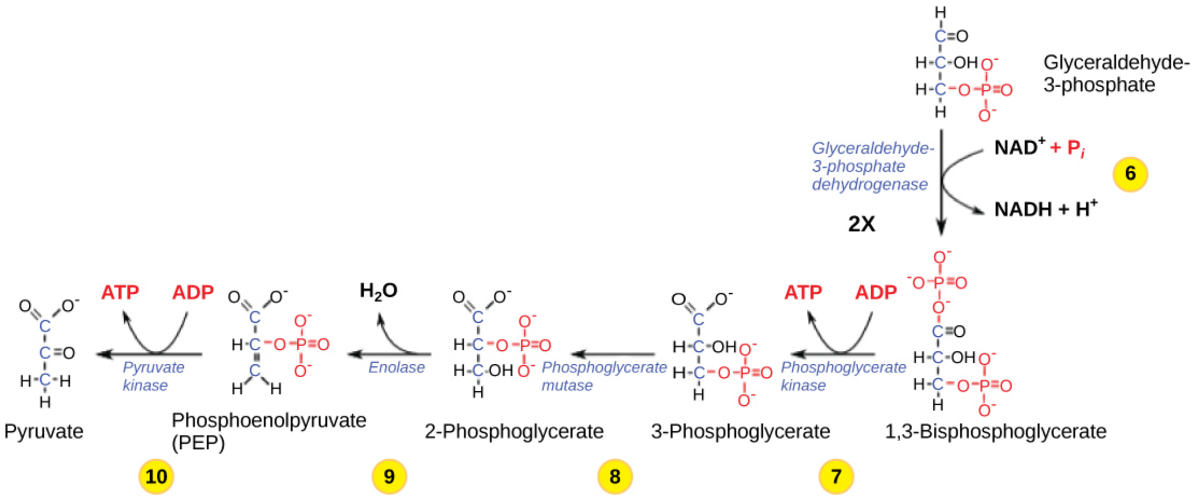


The first half of glycolysis uses two ATP molecules in the phosphorylation of glucose, which is then split into two three-carbon molecules.

Second Half of Glycolysis (Energy-Releasing Steps)

So far, glycolysis has cost the cell two ATP molecules and produced two small, three-carbon sugar molecules. Both of these molecules will proceed through the second half of the pathway, and sufficient energy will be extracted to pay back the two ATP molecules used as an initial investment and produce a profit for the cell of two additional ATP molecules and two even higher-energy NADH molecules.

Step 6. The sixth step in glycolysis ([link]) oxidizes the sugar (glyceraldehyde-3-phosphate), extracting high-energy electrons, which are picked up by the electron carrier NAD^+, producing NADH. The sugar is then phosphorylated by the addition of a second phosphate group, producing 1,3-bisphosphoglycerate. Note that the second phosphate group does not require another ATP molecule.



The second half of glycolysis involves phosphorylation without ATP investment (step 6) and produces two NADH and four ATP molecules per glucose.

Here again is a potential limiting factor for this pathway. The continuation of the reaction depends upon the availability of the oxidized form of the electron carrier, NAD^+. Thus, NADH must be continuously oxidized back into NAD^+ in order to keep this step going. If NAD^+ is not available, the second half of glycolysis slows down or stops. If oxygen is available in the system, the NADH will be oxidized readily, though indirectly, and the high-energy electrons from the hydrogen released in this process will be used to produce ATP. In an environment without oxygen, an alternate pathway (fermentation) can provide the oxidation of NADH to NAD^+.

Step 7. In the seventh step, catalyzed by phosphoglycerate kinase (an enzyme named for the reverse reaction), 1,3-bisphosphoglycerate donates a high-energy phosphate to ADP, forming one molecule of ATP. (This is an example of substrate-level phosphorylation.) A carbonyl group on the 1,3-bisphosphoglycerate is oxidized to a carboxyl group, and 3-phosphoglycerate is formed.

Step 8. In the eighth step, the remaining phosphate group in 3-phosphoglycerate moves from the third carbon to the second carbon, producing 2-phosphoglycerate (an isomer of 3-phosphoglycerate). The enzyme catalyzing this step is a mutase (isomerase).

Step 9. Enolase catalyzes the ninth step. This enzyme causes 2-phosphoglycerate to lose water from its structure; this is a dehydration reaction, resulting in the formation of a double bond that increases the potential energy in the remaining phosphate bond and produces phosphoenolpyruvate (PEP).

Step 10. The last step in glycolysis is catalyzed by the enzyme pyruvate kinase (the enzyme in this case is named for the reverse reaction of pyruvate's conversion into PEP) and results in the production of a second ATP molecule by substrate-level phosphorylation and the compound pyruvic acid (or its salt form, pyruvate). Many enzymes in enzymatic pathways are named for the reverse reactions, since the enzyme can catalyze both forward and reverse reactions (these may have been described initially by the reverse reaction that takes place in vitro, under non-physiological conditions).

Note:
Link to Learning

Gain a better understanding of the breakdown of glucose by glycolysis by visiting this site to see the process in action.

Outcomes of Glycolysis

Glycolysis starts with glucose and ends with two pyruvate molecules, a total of four ATP molecules and two molecules of NADH. Two ATP molecules were used in the first half of the pathway to prepare the six-carbon ring for cleavage, so the cell has a net gain of two ATP molecules and 2 NADH molecules for its use. If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose. Mature mammalian red blood cells are not capable of **aerobic respiration**—the process in which organisms convert energy in the presence of oxygen—and glycolysis is their sole source of ATP. If glycolysis is interrupted, these cells lose their ability to maintain their sodium-potassium pumps, and eventually, they die.

The last step in glycolysis will not occur if pyruvate kinase, the enzyme that catalyzes the formation of pyruvate, is not available in sufficient quantities. In this situation, the entire glycolysis pathway will proceed, but only two ATP molecules will be made in the second half. Thus, pyruvate kinase is a rate-limiting enzyme for glycolysis.

Section Summary

Glycolysis is the first pathway used in the breakdown of glucose to extract energy. It was probably one of the earliest metabolic pathways to evolve and is used by nearly all of the organisms on earth. Glycolysis consists of two parts: The first part prepares the six-carbon ring of glucose for cleavage into two three-carbon sugars. ATP is invested in the process during this half to energize the separation. The second half of glycolysis extracts ATP and high-energy electrons from hydrogen atoms and attaches them to NAD^+. Two ATP molecules are invested in the first half and four ATP molecules are formed by substrate phosphorylation during the second half. This produces a net gain of two ATP and two NADH molecules for the cell.

Review Questions

Exercise:

Problem: During the second half of glycolysis, what occurs?

a. ATP is used up.
b. Fructose is split in two.
c. ATP is made.
d. Glucose becomes fructose.

Solution:

C

Free Response

Exercise:

Problem:

Nearly all organisms on earth carry out some form of glycolysis. How does that fact support or not support the assertion that glycolysis is one of the oldest metabolic pathways?

Solution:

If glycolysis evolved relatively late, it likely would not be as universal in organisms as it is. It probably evolved in very primitive organisms and persisted, with the addition of other pathways of carbohydrate metabolism that evolved later.

Exercise:

Problem:

Red blood cells do not perform aerobic respiration, but they do perform glycolysis. Why do all cells need an energy source, and what would happen if glycolysis were blocked in a red blood cell?

Solution:

All cells must consume energy to carry out basic functions, such as pumping ions across membranes. A red blood cell would lose its membrane potential if glycolysis were blocked, and it would eventually die.

Glossary

aerobic respiration
process in which organisms convert energy in the presence of oxygen

anaerobic
process that does not use oxygen

glycolysis
process of breaking glucose into two three-carbon molecules with the production of ATP and NADH

isomerase
enzyme that converts a molecule into its isomer

pyruvate
three-carbon sugar that can be decarboxylated and oxidized to make acetyl CoA, which enters the citric acid cycle under aerobic conditions; the end product of glycolysis

Oxidation of Pyruvate and the Citric Acid Cycle
By the end of this section, you will be able to:

- Explain how a circular pathway, such as the citric acid cycle, fundamentally differs from a linear pathway, such as glycolysis
- Describe how pyruvate, the product of glycolysis, is prepared for entry into the citric acid cycle

If oxygen is available, aerobic respiration will go forward. In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into mitochondria, which are the sites of cellular respiration. There, pyruvate will be transformed into an acetyl group that will be picked up and activated by a carrier compound called coenzyme A (CoA). The resulting compound is called **acetyl CoA**. CoA is made from vitamin B5, pantothenic acid. Acetyl CoA can be used in a variety of ways by the cell, but its major function is to deliver the acetyl group derived from pyruvate to the next stage of the pathway in glucose catabolism.

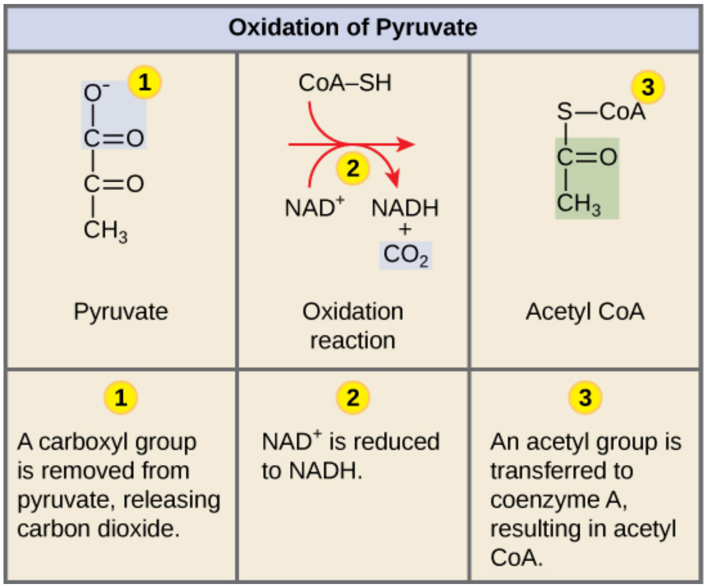
Breakdown of Pyruvate

In order for pyruvate, the product of glycolysis, to enter the next pathway, it must undergo several changes. The conversion is a three-step process ([link]).

Step 1. A carboxyl group is removed from pyruvate, releasing a molecule of carbon dioxide into the surrounding medium. The result of this step is a two-carbon hydroxyethyl group bound to the enzyme (pyruvate dehydrogenase). This is the first of the six carbons from the original glucose molecule to be removed. This step proceeds twice (remember: there are *two* pyruvate molecules produced at the end of glycolsis) for every molecule of glucose metabolized; thus, two of the six carbons will have been removed at the end of both steps.

Step 2. The hydroxyethyl group is oxidized to an acetyl group, and the electrons are picked up by NAD^+, forming NADH. The high-energy electrons from NADH will be used later to generate ATP.

Step 3. The enzyme-bound acetyl group is transferred to CoA, producing a molecule of acetyl CoA.



A carboxyl group is removed from pyruvate, releasing carbon dioxide.

NAD^+ is reduced to NADH.

An acetyl group is transferred to coenzyme A, resulting in acetyl CoA.

Upon entering the mitochondrial matrix, a multi-enzyme complex converts pyruvate into acetyl CoA. In the process, carbon dioxide is released and one molecule of NADH is formed.

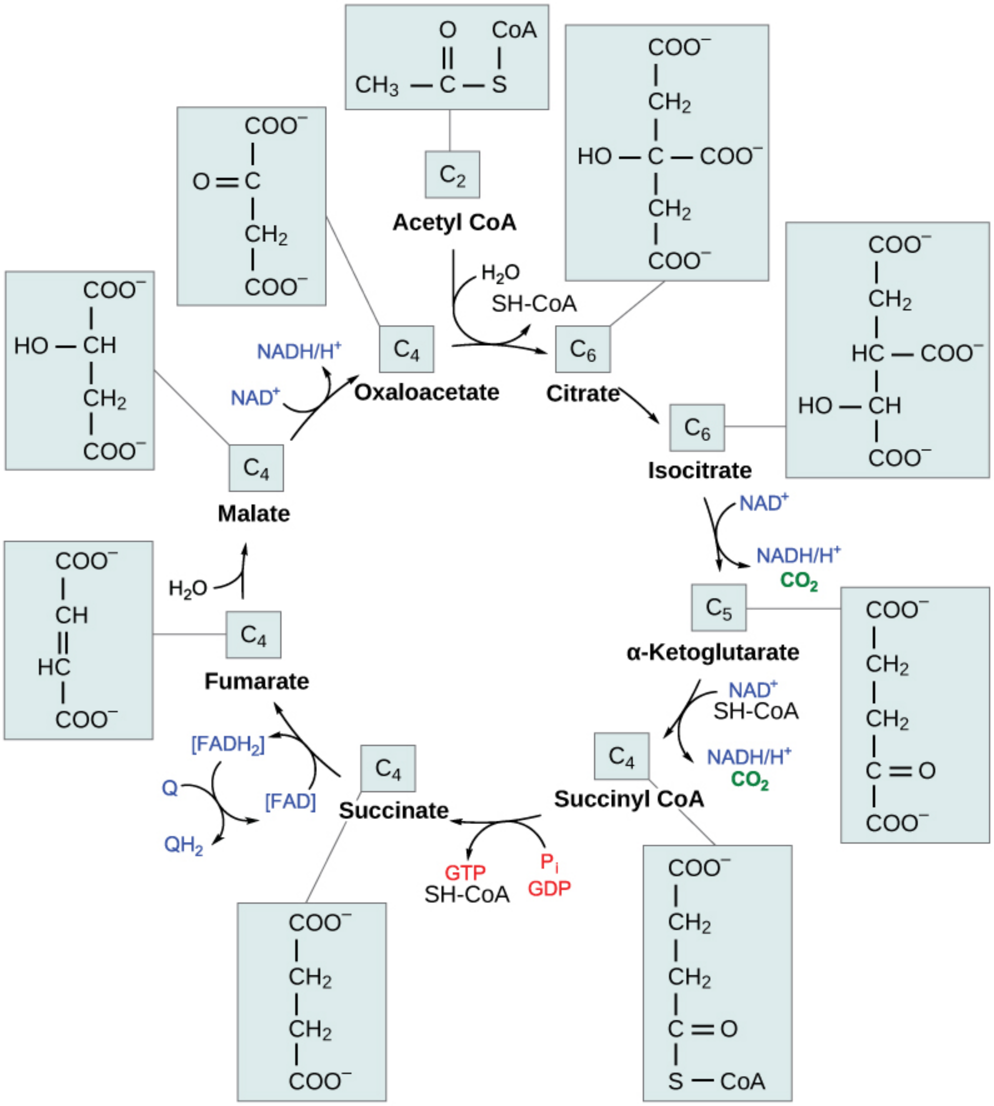
Note that during the second stage of glucose metabolism, whenever a carbon atom is removed, it is bound to two oxygen atoms, producing carbon dioxide, one of the major end products of cellular respiration.

Acetyl CoA to CO_2

In the presence of oxygen, acetyl CoA delivers its acetyl group to a four-carbon molecule, oxaloacetate, to form citrate, a six-carbon molecule with three carboxyl groups; this pathway will harvest the remainder of the extractable energy from what began as a glucose molecule. This single pathway is called by different names: the **citric acid cycle** (for the first intermediate formed—citric acid, or citrate—when acetate joins to the oxaloacetate), the **TCA cycle** (since citric acid or citrate and isocitrate are tricarboxylic acids), and the **Krebs cycle**, after Hans Krebs, who first identified the steps in the pathway in the 1930s in pigeon flight muscles.

Citric Acid Cycle

Like the conversion of pyruvate to acetyl CoA, the citric acid cycle takes place in the matrix of mitochondria. Almost all of the enzymes of the citric acid cycle are soluble, with the single exception of the enzyme succinate dehydrogenase, which is embedded in the inner membrane of the mitochondrion. Unlike glycolysis, the citric acid cycle is a closed loop: The last part of the pathway regenerates the compound used in the first step. The eight steps of the cycle are a series of redox, dehydration, hydration, and decarboxylation reactions that produce two carbon dioxide molecules, one GTP/ATP, and reduced forms of NADH and $FADH_2$ ([link]). This is considered an aerobic pathway because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the system, which will use oxygen. If this transfer does not occur, the oxidation steps of the citric acid cycle also do not occur. Note that the citric acid cycle produces very little ATP directly and does not directly consume oxygen.



In the citric acid cycle, the acetyl group from acetyl CoA is attached to a four-carbon oxaloacetate molecule to form a six-carbon citrate molecule. Through a series of steps, citrate is oxidized, releasing two carbon dioxide molecules for each acetyl group fed into the cycle. In the process, three NAD^+ molecules are reduced to NADH, one FAD molecule is reduced to $FADH_2$, and one ATP or GTP (depending on the cell type) is produced (by substrate-level phosphorylation). Because the final product of the citric acid cycle is also the first reactant, the cycle runs continuously in the presence of sufficient reactants. (credit: modification of work by "Yikrazuul"/Wikimedia Commons)

Steps in the Citric Acid Cycle

Step 1. Prior to the start of the first step, a transitional phase occurs during which pyruvic acid is converted to acetyl CoA. Then, the first step of the cycle begins: This is a condensation step, combining the two-carbon acetyl group with a four-carbon oxaloacetate molecule to form a six-carbon molecule of citrate. CoA is bound to a sulfhydryl group (-SH) and diffuses away to eventually combine with another acetyl group. This step is irreversible because it is highly exergonic. The rate of this reaction is controlled by negative feedback and the amount of ATP available. If ATP levels increase, the rate of this reaction decreases. If ATP is in short supply, the rate increases.

Step 2. In step two, citrate loses one water molecule and gains another as citrate is converted into its isomer, isocitrate.

Step 3. In step three, isocitrate is oxidized, producing a five-carbon molecule, α-ketoglutarate, together with a molecule of CO_2 and two electrons, which reduce NAD^+ to NADH. This step is also regulated by negative feedback from ATP and NADH, and a positive effect of ADP.

Steps 3 and 4. Steps three and four are both oxidation and decarboxylation steps, which release electrons that reduce NAD^+ to NADH and release carboxyl groups that form CO_2 molecules. α-Ketoglutarate is the product of step three, and a succinyl group is the product of step four. CoA binds the succinyl group to form succinyl CoA. The enzyme that catalyzes step four is regulated by feedback inhibition of ATP, succinyl CoA, and NADH.

Step 5. In step five, a phosphate group is substituted for coenzyme A, and a high-energy bond is formed. This energy is used in substrate-level phosphorylation (during the conversion of the succinyl group to succinate) to form either guanine triphosphate (GTP) or ATP. There are two forms of the enzyme, called isoenzymes, for this step, depending upon the type of animal tissue in which they are found. One form is found in tissues that use large amounts of ATP, such as heart and skeletal muscle. This form produces ATP. The second form of the enzyme is found in tissues that have a high number of anabolic pathways, such as liver. This form produces GTP. GTP is energetically equivalent to ATP; however, its use is more restricted. In particular, protein synthesis primarily uses GTP.

Step 6. Step six is a dehydration process that converts succinate into fumarate. Two hydrogen atoms are transferred to FAD, producing $FADH_2$. The energy contained in the electrons of these atoms is insufficient to reduce NAD^+ but adequate to reduce FAD. Unlike NADH, this carrier remains attached to the enzyme and transfers the electrons to the electron transport chain directly. This process is made possible by the localization of the enzyme catalyzing this step inside the inner membrane of the mitochondrion.

Step 7. Water is added to fumarate during step seven, and malate is produced. The last step in the citric acid cycle regenerates oxaloacetate by oxidizing malate. Another molecule of NADH is produced in the process.

Note:
Link to Learning

Click through each step of the citric acid cycle here.

Products of the Citric Acid Cycle

Two carbon atoms come into the citric acid cycle from each acetyl group, representing four out of the six carbons of one glucose molecule. Two carbon dioxide molecules are released on each turn of the cycle; however, these do not necessarily contain the most recently added carbon atoms. The two acetyl carbon atoms will eventually be released on later turns of the cycle; thus, all six carbon atoms from the original glucose molecule are eventually incorporated into carbon dioxide. Each turn of the cycle forms three NADH molecules and one $FADH_2$ molecule. These carriers will connect with the last portion of aerobic respiration to produce ATP molecules. One GTP or ATP is also made in each cycle. Several of the intermediate compounds in the citric acid cycle can be used in synthesizing non-essential amino acids; therefore, the cycle is amphibolic (both catabolic and anabolic).

Section Summary

In the presence of oxygen, pyruvate is transformed into an acetyl group attached to a carrier molecule of coenzyme A. The resulting acetyl CoA can enter several pathways, but most often, the acetyl group is delivered to the citric acid cycle for further catabolism. During the conversion of pyruvate into the acetyl group, a molecule of carbon dioxide and two high-energy electrons are removed. The carbon dioxide accounts for two (conversion of two pyruvate molecules) of the six carbons of the original glucose molecule. The electrons are picked up by NAD^+, and the NADH carries the electrons to a later pathway for ATP production. At this point, the glucose molecule that originally entered cellular respiration has been completely oxidized. Chemical potential energy stored within the glucose molecule has been transferred to electron carriers or has been used to synthesize a few ATPs.

The citric acid cycle is a series of redox and decarboxylation reactions that remove high-energy electrons and carbon dioxide. The electrons temporarily stored in molecules of NADH and $FADH_2$ are used to generate ATP in a subsequent pathway. One molecule of either GTP or ATP is produced by substrate-level phosphorylation on each turn of the cycle. There is no comparison of the cyclic pathway with a linear one.

Review Questions

Exercise:

Problem:

What is removed from pyruvate during its conversion into an acetyl group?

a. oxygen
b. ATP
c. B vitamin
d. carbon dioxide

Solution:

D

Exercise:

Problem: What do the electrons added to NAD^+ do?

a. They become part of a fermentation pathway.
b. They go to another pathway for ATP production.
c. They energize the entry of the acetyl group into the citric acid cycle.
d. They are converted to NADP.

Solution:

B

Exercise:

Problem:GTP or ATP is produced during the conversion of ________.

a. isocitrate into α-ketoglutarate

b. succinyl CoA into succinate
c. fumarate into malate
d. malate into oxaloacetate

Solution:

B

Exercise:

Problem:

How many NADH molecules are produced on each turn of the citric acid cycle?

a. one
b. two
c. three
d. four

Solution:

C

Free Response

Exercise:

Problem:

What is the primary difference between a circular pathway and a linear pathway?

Solution:

In a circular pathway, the final product of the reaction is also the initial reactant. The pathway is self-perpetuating, as long as any of the intermediates of the pathway are supplied. Circular pathways are able to accommodate multiple entry and exit points, thus being particularly well suited for amphibolic pathways. In a linear pathway, one trip through the pathway completes the pathway, and a second trip would be an independent event.

Glossary

acetyl CoA
combination of an acetyl group derived from pyruvic acid and coenzyme A, which is made from pantothenic acid (a B-group vitamin)

citric acid cycle
(also, Krebs cycle) series of enzyme-catalyzed chemical reactions of central importance in all living cells

Krebs cycle
(also, citric acid cycle) alternate name for the citric acid cycle, named after Hans Krebs who first identified the steps in the pathway in the 1930s in pigeon flight muscles; see citric acid cycle

TCA cycle
(also, citric acid cycle) alternate name for the citric acid cycle, named after the group name for citric acid, tricarboxylic acid (TCA); see citric acid cycle

Oxidative Phosphorylation

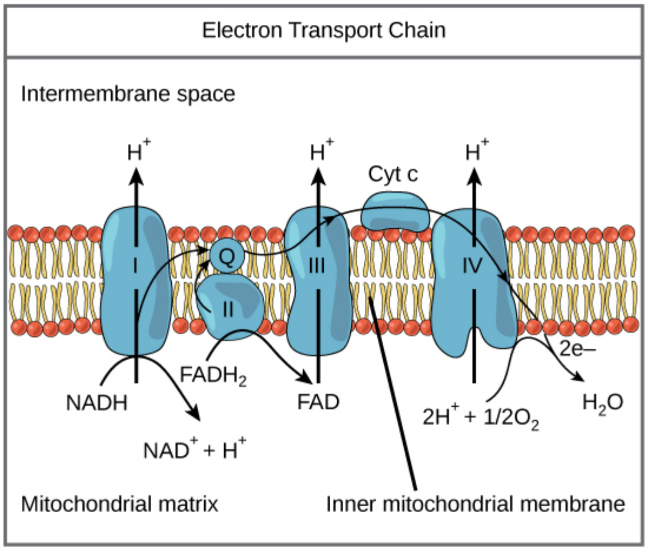
By the end of this section, you will be able to:

- Describe how electrons move through the electron transport chain and what happens to their energy levels
- Explain how a proton (H^+) gradient is established and maintained by the electron transport chain

You have just read about two pathways in glucose catabolism—glycolysis and the citric acid cycle—that generate ATP. Most of the ATP generated during the aerobic catabolism of glucose, however, is not generated directly from these pathways. Rather, it is derived from a process that begins with moving electrons through a series of electron transporters that undergo redox reactions. This causes hydrogen ions to accumulate within the matrix space. Therefore, a concentration gradient forms in which hydrogen ions diffuse out of the matrix space by passing through ATP synthase. The current of hydrogen ions powers the catalytic action of ATP synthase, which phosphorylates ADP, producing ATP.

Electron Transport Chain

The electron transport chain ([link]) is the last component of aerobic respiration and is the only part of glucose metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plants; in animals, it enters the body through the respiratory system. Electron transport is a series of redox reactions that resemble a relay race or bucket brigade in that electrons are passed rapidly from one component to the next, to the endpoint of the chain where the electrons reduce molecular oxygen, producing water. There are four complexes composed of proteins, labeled I through IV in [link], and the aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called the electron transport chain. The electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes and the plasma membrane of prokaryotes.



The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and $FADH_2$ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water.

Complex I

To start, two electrons are carried to the first complex aboard NADH. This complex, labeled I, is composed of flavin mononucleotide (FMN) and an iron-sulfur (Fe-S)-containing protein. FMN, which is derived from vitamin B_2, also called riboflavin, is one of several prosthetic groups or co-factors in the electron transport chain. A **prosthetic group** is a non-protein molecule required for the activity of a protein. Prosthetic groups are organic or inorganic, non-peptide molecules bound to a protein that facilitate its function; prosthetic groups include co-enzymes, which are the prosthetic groups of enzymes. The enzyme in complex I is NADH dehydrogenase and is a very large protein, containing 45 amino acid chains. Complex I can pump four hydrogen ions across the membrane from the matrix into the intermembrane space, and it is in this way that the hydrogen ion gradient is established and maintained between the two compartments separated by the inner mitochondrial membrane.

Q and Complex II

Complex II directly receives $FADH_2$, which does not pass through complex I. The compound connecting the first and second complexes to the third is **ubiquinone** (Q). The Q molecule is lipid soluble and freely moves through the hydrophobic core of the membrane. Once it is reduced, (QH_2), ubiquinone delivers its electrons to the next complex in the electron transport chain. Q receives the electrons derived from NADH from complex I and the electrons derived from $FADH_2$ from complex II, including succinate dehydrogenase. This enzyme and $FADH_2$ form a small complex that delivers electrons directly to the electron transport chain, bypassing the first complex. Since these electrons bypass and thus do not energize the proton pump in the first complex, fewer ATP molecules are made from the $FADH_2$ electrons. The number of ATP molecules ultimately obtained is directly proportional to the number of protons pumped across the inner mitochondrial membrane.

Complex III

The third complex is composed of cytochrome b, another Fe-S protein, Rieske center (2Fe-2S center), and cytochrome c proteins; this complex is also called cytochrome oxidoreductase. Cytochrome proteins have a prosthetic group of heme. The heme molecule is similar to the heme in hemoglobin, but it carries electrons, not oxygen. As a result, the iron ion at its core is reduced and oxidized as it passes the electrons, fluctuating between different oxidation states: Fe^{++} (reduced) and Fe^{+++} (oxidized). The heme molecules in the cytochromes have slightly different characteristics due to the effects of the different proteins binding them, giving slightly different characteristics to each complex. Complex III pumps protons through the membrane and passes its electrons to cytochrome c for transport to the fourth complex of proteins and enzymes (cytochrome c is the acceptor of electrons from Q; however, whereas Q carries pairs of electrons, cytochrome c can accept only one at a time).

Complex IV

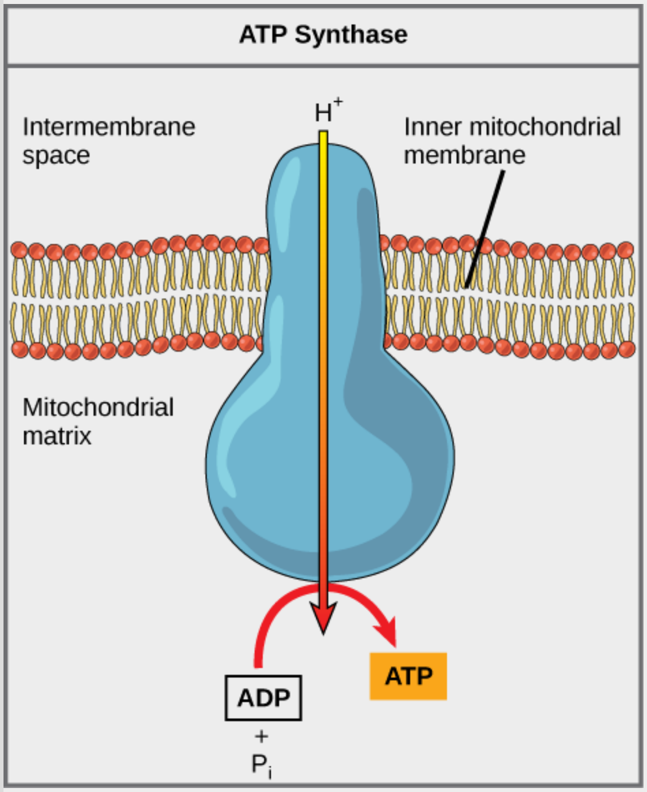
The fourth complex is composed of cytochrome proteins c, a, and a_3. This complex contains two heme groups (one in each of the two cytochromes, a, and a_3) and three copper ions (a pair of Cu_A and one Cu_B in cytochrome a_3). The cytochromes hold an oxygen molecule very tightly between the iron and copper ions until the oxygen is completely reduced. The reduced oxygen then picks up two hydrogen ions from the surrounding medium to make water (H_2O). The removal of the hydrogen ions from the system contributes to the ion gradient used in the process of chemiosmosis.

Chemiosmosis

In chemiosmosis, the free energy from the series of redox reactions just described is used to pump hydrogen ions (protons) across the membrane. The uneven distribution of H^+ ions across the membrane establishes both concentration and electrical gradients (thus, an electrochemical gradient), owing to the hydrogen ions’ positive charge and their aggregation on one side of the membrane.

If the membrane were open to diffusion by the hydrogen ions, the ions would tend to diffuse back across into the matrix, driven by their electrochemical gradient. Recall that many ions cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, hydrogen ions in the matrix space can only pass through the inner mitochondrial membrane through an integral membrane protein called ATP synthase ([link]). This complex protein acts as a tiny generator, turned by the force of the hydrogen ions diffusing through it, down their electrochemical gradient. The turning of parts of this molecular machine facilitates the addition of a phosphate to ADP, forming ATP, using the potential energy of the hydrogen ion gradient.

Note:
Art Connection



ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (Pi). (Credit: modification of work by Klaus Hoffmeier)

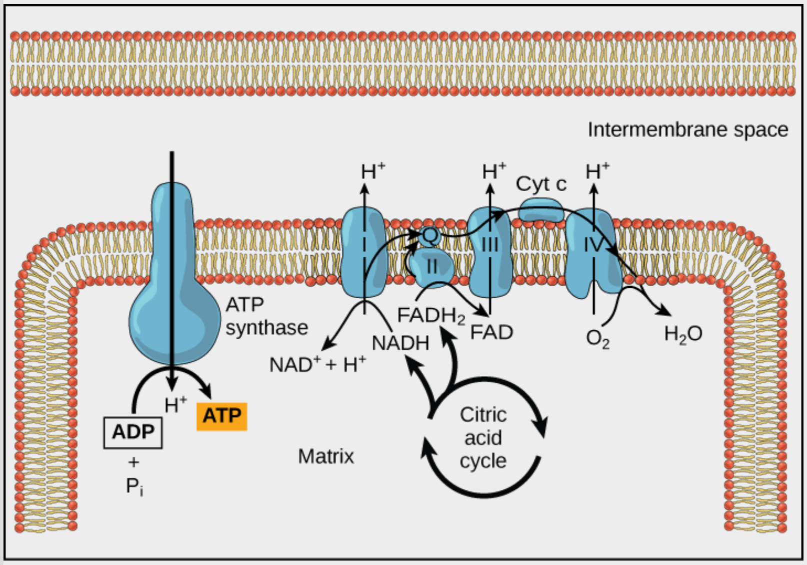
Dinitrophenol (DNP) is an uncoupler that makes the inner mitochondrial membrane leaky to protons. It was used until 1938 as a weight-loss drug.

What effect would you expect DNP to have on the change in pH across the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug?

Chemiosmosis ([link]) is used to generate 90 percent of the ATP made during aerobic glucose catabolism; it is also the method used in the light reactions of photosynthesis to harness the energy of sunlight in the process of photophosphorylation. Recall that the production of ATP using the process of chemiosmosis in mitochondria is called oxidative phosphorylation. The overall result of these reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end of the pathway, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen attract hydrogen ions (protons) from the surrounding medium, and water is formed.

Note:

Art Connection



In oxidative phosphorylation, the pH gradient formed by the electron transport chain is used by ATP synthase to form ATP.

Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What effect would cyanide have on ATP synthesis?

ATP Yield

The number of ATP molecules generated from the catabolism of glucose varies. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Another source of variance stems from the shuttle of electrons across the membranes of the mitochondria. (The NADH generated from glycolysis cannot easily enter mitochondria.) Thus, electrons are picked up on the inside of mitochondria by either NAD^+ or FAD^+. As you have learned earlier, these FAD^+ molecules can transport fewer ions; consequently, fewer ATP molecules are generated when FAD^+ acts as a carrier. NAD^+ is used as the electron transporter in the liver and FAD^+ acts in the brain.

Another factor that affects the yield of ATP molecules generated from glucose is the fact that intermediate compounds in these pathways are used for other purposes. Glucose catabolism connects with the pathways that build or break down all other biochemical compounds in cells, and the result is somewhat messier than the ideal situations described thus far. For example, sugars other than glucose are fed into the glycolytic pathway for energy extraction. Moreover, the five-carbon sugars that form nucleic acids are made from intermediates in glycolysis. Certain nonessential amino acids can be made from intermediates of both glycolysis and the citric acid cycle. Lipids, such as cholesterol and triglycerides, are also made from intermediates in these pathways, and both amino acids and triglycerides are broken down for energy through these pathways. Overall, in living systems, these pathways of glucose catabolism extract about 34 percent of the energy contained in glucose.

Section Summary

The electron transport chain is the portion of aerobic respiration that uses free oxygen as the final electron acceptor of the electrons removed from the intermediate compounds in glucose catabolism. The electron transport chain is composed of four large, multiprotein complexes embedded in the inner mitochondrial membrane and two small diffusible electron carriers shuttling electrons between them. The electrons are passed through a series of redox reactions, with a small amount of free energy used at three points to transport hydrogen ions across a membrane. This process contributes to the gradient used in chemiosmosis. The electrons passing through the electron transport chain gradually lose energy, High-energy electrons donated to the chain by either NADH or $FADH_2$ complete the chain, as low-energy electrons reduce oxygen molecules and form water. The level of free energy of the electrons drops from about 60 kcal/mol in NADH or 45 kcal/mol in $FADH_2$ to about 0 kcal/mol in water. The end products of the electron transport chain are water and ATP. A number of intermediate compounds of the citric acid cycle can be diverted into the anabolism of other biochemical molecules, such as nonessential amino acids, sugars, and lipids. These same molecules can serve as energy sources for the glucose pathways.

Art Connections

Exercise:

Problem:

[link] Dinitrophenol (DNP) is an uncoupler that makes the inner mitochondrial membrane leaky to protons. It was used until 1938 as a weight-loss drug. What effect would you expect DNP to have on the change in pH across the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug?

Solution:

[link] After DNP poisoning, the electron transport chain can no longer form a proton gradient, and ATP synthase can no longer make ATP. DNP is an effective diet drug because it uncouples ATP synthesis; in

other words, after taking it, a person obtains less energy out of the food he or she eats. Interestingly, one of the worst side effects of this drug is hyperthermia, or overheating of the body. Since ATP cannot be formed, the energy from electron transport is lost as heat.

Exercise:

Problem:

[link] Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What effect would cyanide have on ATP synthesis?

Solution:

[link] After cyanide poisoning, the electron transport chain can no longer pump electrons into the intermembrane space. The pH of the intermembrane space would increase, the pH gradient would decrease, and ATP synthesis would stop.

Review Questions

Exercise:

Problem: What compound receives electrons from NADH?

a. FMN
b. ubiquinone
c. cytochrome c_1
d. oxygen

Solution:

A

Exercise:

Problem: Chemiosmosis involves ________.

a. the movement of electrons across the cell membrane
b. the movement of hydrogen atoms across a mitochondrial membrane
c. the movement of hydrogen ions across a mitochondrial membrane
d. the movement of glucose through the cell membrane

Solution:

C

Free Response

Exercise:

Problem:

How do the roles of ubiquinone and cytochrome c differ from the other components of the electron transport chain?

Solution:

Q and cytochrome c are transport molecules. Their function does not result directly in ATP synthesis in that they are not pumps. Moreover, Q is the only component of the electron transport chain that is not a protein. Ubiquinone and cytochrome c are small, mobile, electron carriers, whereas the other components of the electron transport chain are large complexes anchored in the inner mitochondrial membrane.

Exercise:

Problem:

What accounts for the different number of ATP molecules that are formed through cellular respiration?

Solution:

Few tissues except muscle produce the maximum possible amount of ATP from nutrients. The intermediates are used to produce needed amino acids, fatty acids, cholesterol, and sugars for nucleic acids. When NADH is transported from the cytoplasm to the mitochondria, an active transport mechanism is used, which decreases the amount of ATP that can be made. The electron transport chain differs in composition between species, so different organisms will make different amounts of ATP using their electron transport chains.

Glossary

ATP synthase

(also, F1F0 ATP synthase) membrane-embedded protein complex that adds a phosphate to ADP with energy from protons diffusing through it

prosthetic group

(also, prosthetic cofactor) molecule bound to a protein that facilitates the function of the protein

ubiquinone

soluble electron transporter in the electron transport chain that connects the first or second complex to the third

Metabolism without Oxygen

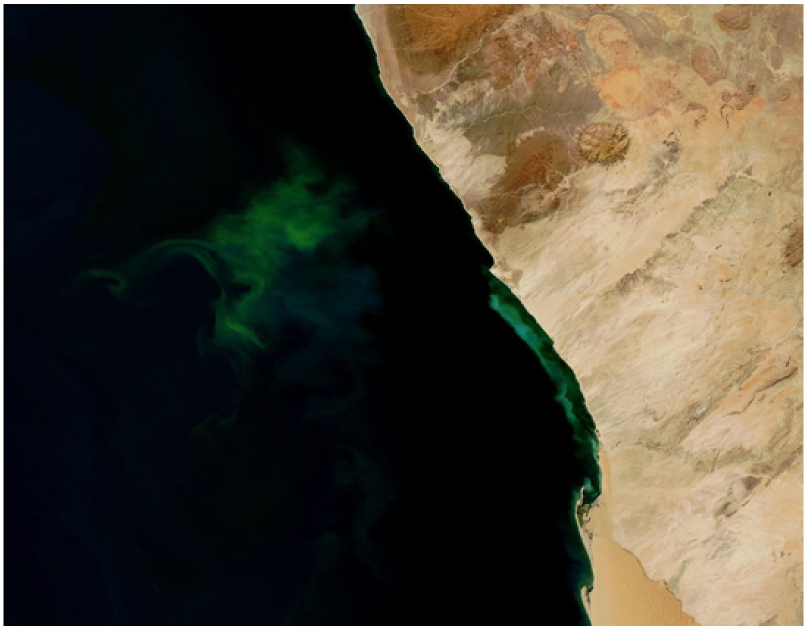
By the end of this section, you will be able to:

- Discuss the fundamental difference between anaerobic cellular respiration and fermentation
- Describe the type of fermentation that readily occurs in animal cells and the conditions that initiate that fermentation

In aerobic respiration, the final electron acceptor is an oxygen molecule, O_2. If aerobic respiration occurs, then ATP will be produced using the energy of high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If aerobic respiration does not occur, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for the glycolytic pathway to continue. How is this done? Some living systems use an organic molecule as the final electron acceptor. Processes that use an organic molecule to regenerate NAD^+ from NADH are collectively referred to as **fermentation**. In contrast, some living systems use an inorganic molecule as a final electron acceptor. Both methods are called **anaerobic cellular respiration** in which organisms convert energy for their use in the absence of oxygen.

Anaerobic Cellular Respiration

Certain prokaryotes, including some species of bacteria and Archaea, use anaerobic respiration. For example, the group of Archaea called methanogens reduces carbon dioxide to methane to oxidize NADH. These microorganisms are found in soil and in the digestive tracts of ruminants, such as cows and sheep. Similarly, sulfate-reducing bacteria and Archaea, most of which are anaerobic ([link]), reduce sulfate to hydrogen sulfide to regenerate NAD^+ from NADH.



The green color seen in these coastal waters is from an eruption of hydrogen sulfide-producing bacteria. These anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas as they decompose algae in the water. (credit: modification of work by NASA/Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC, Visible Earth Catalog of NASA images)

Note:
Link to Learning

Visit this site to see anaerobic cellular respiration in action.

Lactic Acid Fermentation

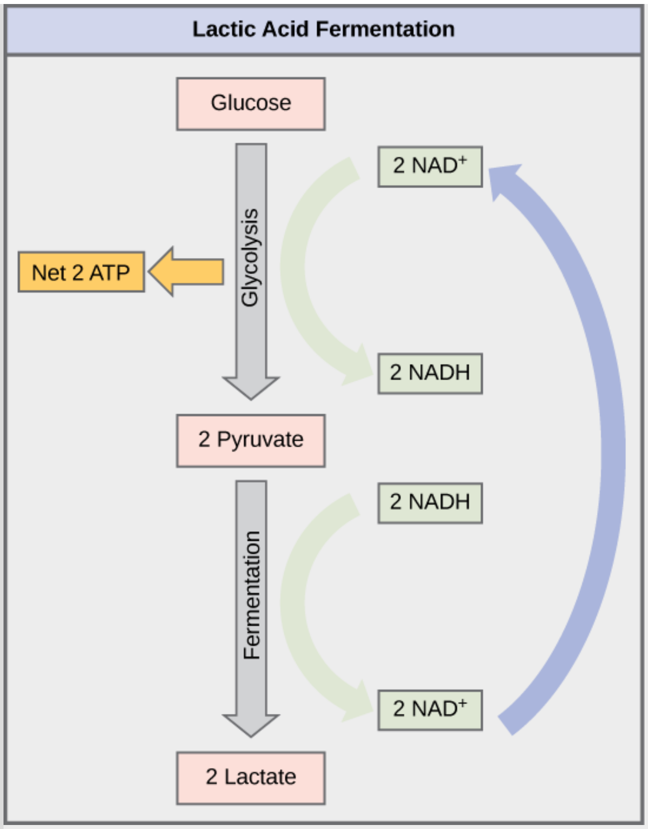
The fermentation method used by animals and certain bacteria, like those in yogurt, is lactic acid fermentation ([link]). This type of fermentation is used routinely in mammalian red blood cells and in skeletal muscle that has an insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid accumulation must be removed by the blood circulation and the lactate brought to the liver for further metabolism. The chemical reactions of lactic acid fermentation are the following:

Equation:

$$\text{Pyruvic acid} + \text{NADH} \leftrightarrow \text{lactic acid} + \text{NAD}^{+}$$

The enzyme used in this reaction is lactate dehydrogenase (LDH). The reaction can proceed in either direction, but the reaction from left to right is inhibited by acidic conditions. Such lactic acid accumulation was once believed to cause muscle stiffness, fatigue, and soreness, although more recent research disputes this hypothesis. Once the lactic acid has been removed from the muscle and circulated to the liver, it can be reconverted into pyruvic acid and further catabolized for energy.

Note:
Art Connection



Lactic acid fermentation is common in muscle cells that have run out of oxygen.

Tremetol, a metabolic poison found in the white snake root plant, prevents the metabolism of lactate. When cows eat this plant, it is concentrated in the milk they produce. Humans who consume the milk become ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation ([link]) that produces ethanol, an alcohol. The first chemical reaction of alcohol fermentation is the following (CO_2 does not participate in the second reaction):

Equation:

$$\text{Pyruvic acid} \rightarrow CO_2 + \text{acetaldehyde} + \text{NADH} \rightarrow \text{ethanol} + \text{NAD}^+$$

The first reaction is catalyzed by pyruvate decarboxylase, a cytoplasmic enzyme, with a coenzyme of thiamine pyrophosphate (TPP, derived from vitamin B_1 and also called thiamine). A carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the size of the molecule by one carbon, making acetaldehyde. The second reaction is catalyzed by alcohol dehydrogenase to oxidize NADH to NAD^+ and reduce acetaldehyde to ethanol. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages. Ethanol tolerance of yeast is variable, ranging from about 5 percent to 21 percent, depending on the yeast strain and environmental conditions.

Fermentation of grape juice into wine produces CO_2 as a byproduct. Fermentation tanks have valves so that the pressure inside the tanks

created by the carbon dioxide
produced can be released.

Other Types of Fermentation

Other fermentation methods occur in bacteria. Many prokaryotes are facultatively anaerobic. This means that they can switch between aerobic respiration and fermentation, depending on the availability of oxygen. Certain prokaryotes, like *Clostridia*, are obligate anaerobes. Obligate anaerobes live and grow in the absence of molecular oxygen. Oxygen is a poison to these microorganisms and kills them on exposure. It should be noted that all forms of fermentation, except lactic acid fermentation, produce gas. The production of particular types of gas is used as an indicator of the fermentation of specific carbohydrates, which plays a role in the laboratory identification of the bacteria. Various methods of fermentation are used by assorted organisms to ensure an adequate supply of NAD^+ for the sixth step in glycolysis. Without these pathways, that step would not occur and no ATP would be harvested from the breakdown of glucose.

Section Summary

If NADH cannot be oxidized through aerobic respiration, another electron acceptor is used. Most organisms will use some form of fermentation to accomplish the regeneration of NAD^+, ensuring the continuation of glycolysis. The regeneration of NAD^+ in fermentation is not accompanied by ATP production; therefore, the potential of NADH to produce ATP using an electron transport chain is not utilized.

Art Connections

Exercise:

Problem:

[link] Tremetol, a metabolic poison found in the white snake root plant, prevents the metabolism of lactate. When cows eat this plant, it is concentrated in the milk they produce. Humans who consume the milk become ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Solution:

[link] The illness is caused by lactate accumulation. Lactate levels rise after exercise, making the symptoms worse. Milk sickness is rare today, but was common in the Midwestern United States in the early 1800s.

Review Questions

Exercise:

Problem:

Which of the following fermentation methods can occur in animal skeletal muscles?

a. lactic acid fermentation
b. alcohol fermentation
c. mixed acid fermentation
d. propionic fermentation

Solution:

A

Free Response

Exercise:

Problem:

What is the primary difference between fermentation and anaerobic respiration?

Solution:

Fermentation uses glycolysis only. Anaerobic respiration uses all three parts of cellular respiration, including the parts in the mitochondria like the citric acid cycle and electron transport; it also uses a different final electron acceptor instead of oxygen gas.

Glossary

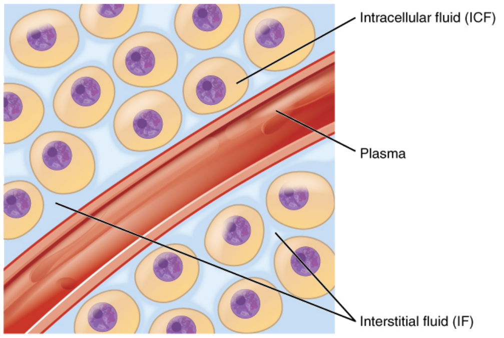
anaerobic cellular respiration
: process in which organisms convert energy for their use in the absence of oxygen

fermentation
: process of regenerating NAD^+ with either an inorganic or organic compound serving as the final electron acceptor, occurs in the absence; occurs in the absence of oxygen

Introduction to Membrane Transport

An introduction to membrane transport that covers cell and tissue membranes, body water compartments, and the basic limiting factors that membranes use to selectively allow passage of ions and molecules.

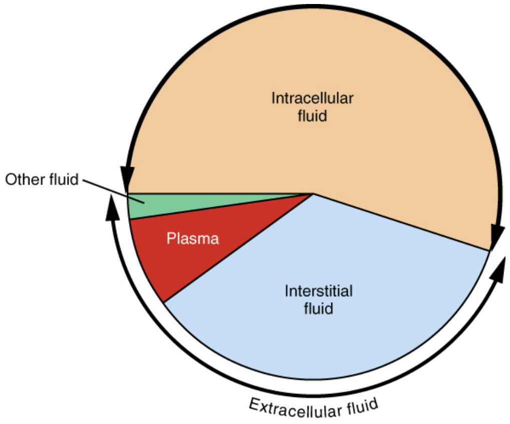
Life Happens in water. Even though the outside of our bodies exist in a relatively dry environment, on the inside our cells are surrounded by water. Different *bags* (compartments) of water have different compositions allowing them to carry out different functions. At the highest level we can separate these compartments into the **Extracellular Fluid (ECF)**, the water environment shared by all the cells in the body, and the the **Intracellular Fluid (ICF)**, the water environment unique to an individual cell. The ICF is contained and separated by individual cell membranes while the ECF is contained and separated by tissue membranes that consist of tightly joined cell membranes and their constituent cells. ([link])



Three fluid compartments in the body

Extracellular Fluid Compartments

The ECF is divided into three system-wide sub-compartments based on function and the composition of the fluid contained in that sub-compartment. ([link])



Distribution of fluid among the fluid compartments.

Plasma is the fluid component of blood found in the circulatory system. It contains blood cells and cell fragments, large proteins, nutrients, gasses, ions, and wastes. It is the main distribution medium for all the chemicals in the body, moving them from one compartment to the other. As such the fluids in the other compartments contain selective subsets of the components found in plasma.

Interstitial Fluid refers to the fluid immediately surrounding the individual cells. Interstitial means the spaces between. At the mass action level it may seem like interstitial fluid is relatively static. However, it is continually being renewed by fluid diffusion into and out of the capillaries that supply the area. It contains a subset of the constituents found in the plasma from which it is derived. The capillaries restrict the dissolved components in the plasma by size. Anything too large to pass through their filtration slits is retained in the plasma and not found in the interstitial fluid. Cells and the large globular proteins are found in plasma but not in the interstitial fluid because under normal conditions they are too large to pass through the filtration slits.

The interstitial fluid is the medium in direct contact with the cells. The cells selectively move molecules and ions both from the interstitial fluid into the cell and from the cell back into the interstitial fluid. This modifies the interstitial fluid contents basically removing those molecules and ions needed by the cell and adding those the cell discards as waste. At the venous end of the capillaries interstitial fluid is reabsorbed and once again becomes part of the plasma. This cellular modification of the interstitial fluid creates the differences found in the chemical composition between plasma in the arteries and plasma in the veins.

Due to pressure and concentration differences not all the plasma fluid that diffused into the interstitial space is able to be reabsorbed by the circulatory system capillaries. The extra interstitial fluid needs to be removed or it will continue to build up swelling the tissue in the area causing a condition called *edema*. To accomplish this a separate set of very low pressure **lymphatic capillaries** remove the fluid, eventually returning it to the plasma in the closed loop of the circulatory vessels.

Lymph refers to the fluid found in the lymphatic vessels and their specialized tissues and organs. Lymphatic capillaries have larger, specialized, openings than circulatory capillaries. These allow larger proteins and cells to be reabsorbed with the interstitial fluid. Lymph is the fluid where most of the body's immune responses take place. Special lymphatic capillaries found in the small intestines, called *lacteals*, also play an important role in absorbing and transporting lipids into the body. Therefore lymph from this area contains a higher concentration of these lipid materials than is found in interstitial fluid or plasma.

There are two specialized sub-compartments in addition to the three system-wide sub-compartments described above. These are **Cerebral Spinal Fluid (CSF)** that bathes the cells of the central nervous system and **Glomerular Filtrate**. Like interstitial fluid and lymph, both are derived from plasma. Glomerular filtrate is different from the other compartments in that after selective reabsorption and secretion exchanges with the plasma the remaining filtrate and its components, now called urine, are removed from the body as waste rather than returned to the plasma.

Selectivity Mechanisms In Membrane Transport

Unlike extracellular fluid, which is shared by the body as a whole, intracellular fluid composition varies depending on the specific cell and its current activities. The composition depends on the ability of the cell membrane to selectively admit or exclude different substances. Many of the differences between the extracellular fluid compartments also are mediated by intracellular movement of substances through the cells making up the tissue membrane. The type of membrane transport used for a specific molecule depends first on whether the molecule is hydrophilic or hydrophobic.

Hydrophobic molecules: You will recall that the inner portion of the membrane consists of lipid tails and therefore is hydrophobic. While this non-polar layer is what creates the water barrier between the outside and inside of the cell, it also means that other non-polar molecules can readily diffuse across the membrane. Therefore **hydrophobic (non-polar) molecules** are not selectively limited by cell membranes. Lipids and other non-polar molecules can readily diffuse across the cell membrane.

Hydrophilic molecules are polar and therefore are generally excluded from crossing the cell membrane without a channel that allows them through. This is much like having special doors and windows that allow movement of objects and molecules into and out of a room. The characteristics of these channels determine which specific molecules are allowed through, and sometimes also control the direction in which the molecules can move. Cell channels use three basic features of hydrophilic molecules and ions to selectively control which types can move into or out of the cell.

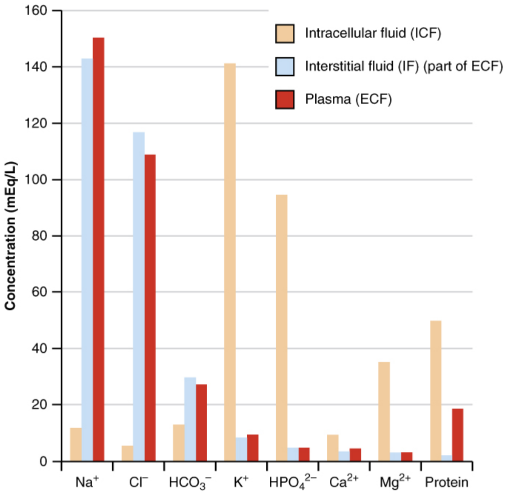
Size: The most generic form of selectivity is to limit the size of the opening in the channel. Only molecules that are smaller than the opening can pass through. While this excludes molecules too large to pass through, any molecule that is smaller can pass through, so the selectivity is somewhat limited.

Electrical Charge: Channels are made of proteins, and folded proteins often have different areas that have neutral, positive, or negative charges. Because opposite electrical charges attract each other and like charges repel each other, the pattern of electrical charges in the channels can further select which molecules can move through the channel.

Molecular Shape: In addition to size and electrical charge patterns the selectivity can be increased by restricting the three-dimensional shape of the molecules that can pass through the channel. These generally are not a fixed shape but act in a manner similar to induced fit. The channel and molecule change shape as they interact and the molecular dance moves the molecule through the channel. These channels can be highly specific in the molecules they allow through the membrane.

Composition Differences Between ECF and ICF

The compositions of the two components of the ECF—plasma and IF—are more similar to each other than either is to the ICF ([link]). Blood plasma has high concentrations of sodium, chloride, bicarbonate, and protein. The IF has high concentrations of sodium, chloride, and bicarbonate, but a relatively lower concentration of protein. In contrast, the ICF has elevated amounts of potassium, phosphate, magnesium, and protein. Overall, the ICF contains high concentrations of potassium and phosphate (HPO_4^{2-}), whereas both plasma and the ECF contain high concentrations of sodium and chloride.



Distribution of common elements in body fluids

Note:

Watch this video to learn more about body fluids, fluid compartments, and electrolytes. When blood volume decreases due to sweating, from what source is water taken in by the blood?

Chapter Review

Your body is mostly water. Body fluids are aqueous solutions with differing concentrations of materials, called solutes. These fluids are compartmentalized from each other by cell and organ membranes. The compartments can be divided into intracellular fluids and extracellular fluids. Intracellular fluids are those found inside cells. Extracellular fluids are further divided by body compartments into **plasma**, found in blood vessels and the heart; **interstitial fluid**, found surrounding the living cells in the body; and **lymph**, found in special very low pressure vessels that carry a small portion of the interstitial fluid and other materials like absorbed lipids to the blood vessels. In addition to these three systemic types of body fluids there are two additional more localized types of body fluids. **cerebral spinal fluid** surrounds the cells of the central nervous system, and **glomerular filtrate** is transformed by the kidneys into urine for excretion from the body.

All of the body fluids originate as plasma and are modified by the cells as they move from one compartment to the next. These cells selectively move molecules and ions through their plasma membranes. Non-polar molecules can diffuse directly through the membrane. Polar molecules and ions can be selectively transferred across the membrane by specialized protein channels. These channels select what can cross the membrane base on size, electrical charge, and shape of the molecule. As a result the chemical composition of the fluids in each of these compartments differ from each other.

Forces That Move Ions and Molecules
By the end of this section, you will be able to:

- Describe the molecular components that make up the cell membrane
- Explain the major features and properties of the cell membrane
- Differentiate between materials that can and cannot diffuse through the lipid bilayer
- Compare and contrast different types of passive transport with active transport, providing examples of each

Transport across the Cell Membrane

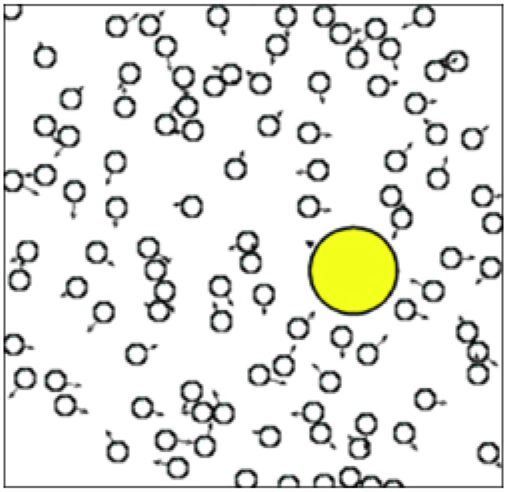
Despite differences in structure and function, all living cells in multicellular organisms have a surrounding cell membrane. As the outer layer of your skin separates your body from its environment, the plasma membrane separates the inner contents of a cell from its exterior environment. This plasma membrane provides a protective barrier around the cell and regulates which materials can pass in or out. One of the great wonders of the plasma membrane is its ability to regulate the concentration of substances inside the cell. These substances include ions such as Ca^{++}, Na^+, K^+, and $Cl^–$; nutrients including sugars, fatty acids, and amino acids; and waste products, particularly carbon dioxide (CO_2), which must leave the cell.

The membrane’s lipid bilayer structure provides the first level of control. The phospholipids are tightly packed together, and the membrane has a hydrophobic interior. This structure causes the membrane to be selectively permeable. A membrane that has **selective permeability** allows only substances meeting certain criteria to pass through it unaided. In the case of the cell membrane, only relatively small, nonpolar materials can move through the lipid bilayer (remember, the lipid tails of the membrane are nonpolar). Some examples of these are other lipids, oxygen and carbon dioxide gases, and alcohol. However, water-soluble materials—like glucose, amino acids, and electrolytes—need some assistance to cross the membrane because they are repelled by the hydrophobic tails of the phospholipid bilayer. All substances that move through the membrane do so by one of two general methods, which are categorized based on whether or not energy is required. **Passive transport** is the movement of substances across the membrane without the expenditure of cellular energy. In contrast, **active transport** is the movement of substances across the membrane using a high energy molecule like adenosine triphosphate (ATP).

Driving Forces

Brownian Motion

Robert Brown, the nineteenth-century British botanist and surgeon, while looking at pollen grains suspended in water under a microscope, noted that the pollen granules appeared to be in a constant state of agitation and also seemed to demonstrate a vivid, oscillatory motion. At first he attributed this motion to activity of the living pollen grains. However, with further research he discovered the same motion occurred with small particles of non-living material as well. The physical phenomena described by Robert Brown have come collectively to be known in his honor by the term Brownian motion. Brownian motion has since been shown to be a simple stochastic process that can be modeled mathematically to characterize the random movements of minute particles immersed in fluids.

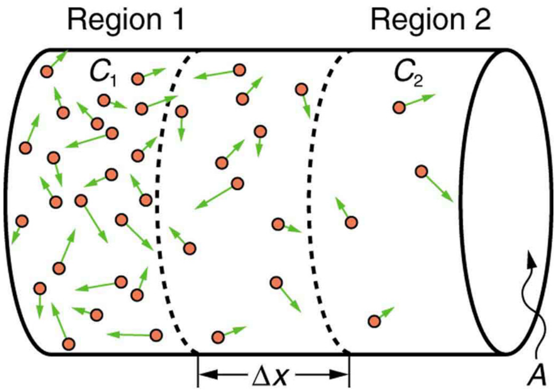


A grain of pollen colliding with water molecules

Brownian motion arises from the agitation of individual molecules by thermal energy. The collective impact of the suspension fluid molecules against the suspended particle yields enough momentum to create movement of the particle in spite of its sometimes exponentially larger size ([link]). The inherently random nature of the motions of the fluid molecules cause the suspended particles to also move randomly. The properties of Brownian motion have been found to be applicable to all diffusion phenomena.

Chemical Gradients

A **concentration gradient** is the difference in the concentration of a substance across a space. **Diffusion** is the movement of molecules from an area of higher concentration to an area of lower concentration. Brownian motion will cause the molecules (or ions) to spread/diffuse from where they are more concentrated to where they are less concentrated until they are equally distributed in that space ([link]). When molecules move in this way, they are said to move *down* their concentration gradient. In living cells the solvent in which the solute particles are dissolved is water. The Brownian motion provides the energy that keeps the particles in constant motion.



Diffusion proceeds from a region of higher concentration to a lower one. The net rate of movement is

proportional to the difference
in concentration.

Note:

Thought Experiment

A simple thought experiment will help to illustrate the process by which Brownian motion powers diffusion. Imagine that everyone in your class is standing close together in the corner of an empty room. Everyone is given one instruction, they have to keep moving, they cannot stop, just like the Brownian motion of suspended molecules. At this point the lights are turned off and the room is completely dark. No one can see anyone else or where they are in the room, but they all keep moving. After 15 minutes everyone is told to stop where they are and the lights are turned on. Remember everyone started standing close together in one corner of the room. When the lights are turned on do you think they will all be in the same corner of the room? If not, how do you think they will be distributed around the room? Why?

Do you think the speed at which everyone is moving (walking slow versus walking fast) would have an effect on how quickly they reach equal distribution? The speed of movement of molecules is directly related to the temperature of the solution. What conclusions can you draw about temperature and diffusion?

Would there be a difference in how quickly equal distribution would be reached if we ran the same experiment twice, but started with twice as many people in the corner in the second trial? What conclusions can you draw about concentration (number of people) in the corner and diffusion?

Note:

Visit this link to see diffusion and how it is propelled by the kinetic energy of molecules in solution. How does temperature affect diffusion rate, and why?

Diffusion is the dominant mechanism by which the exchange of nutrients and waste products occur between the blood and tissue, and between air and blood in the lungs. In the evolutionary process, as organisms became larger, they needed faster methods than net diffusion to transport materials over the larger distances, leading to the development of circulatory systems. However, circulatory systems deliver to and pick up from the interstitial fluid around the cells. Diffusion through the interstitial fluid carries these materials between the cells and the circulatory vessels. Smaller, single-celled organisms still rely totally on diffusion for the removal of waste products and the uptake of nutrients. All cells, including human cells, rely on diffusion to move key materials around the inside of the cell, which exerts a major limitation on the maximum size a cell can be.

Electrical charges and electrical gradients.
As discussed in the chapter on chemical reactions, ions have an electrical charge. So do many biological molecules. These charges are identified as being positive or negative. Ions and molecules with opposite charges are attracted to each other while ions and molecules with the same charge repel (push away from) each other. If you have ever played with magnets you have experienced an analogous phenomenon. Similar to electrical charges, magnetic fields also have two poles, and opposite poles are attracted to each other while like poles repel each other. If you orient two magnets so their *opposite poles* are aligned they are pulled together and held together by the attraction of the opposing forces. However, if you orient the two magnets so

their *like poles* are aligned they resist being pushed together and if you can get them together they will not stay together without continued force.

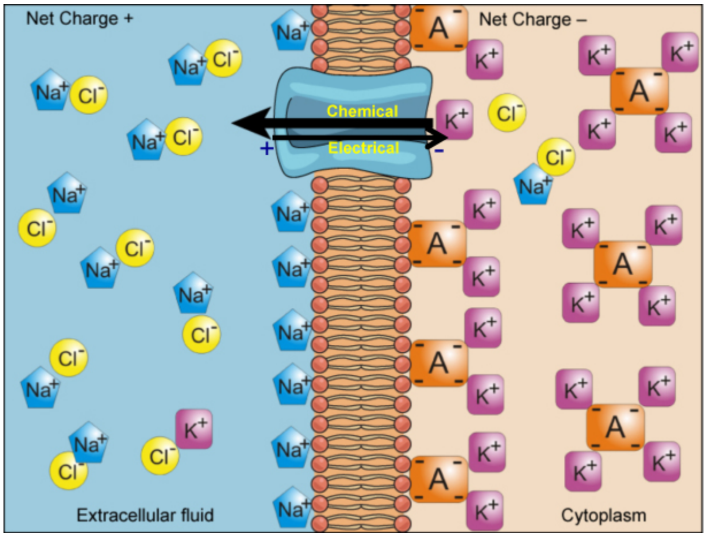
Electrical charges have these same properties. An **electrical gradient**, the difference in the strength of the electrical charges between two separated concentrations of ions or molecules, is measured in **volts**. The greater the voltage (the electrical gradient) the more strongly ions and molecules will be attracted to a concentration with their opposite charge and the more strongly they will be repelled by a concentration with their same charge.

The electrical forces are important in biology because the extracellular fluid is positively charged, while the intracellular fluid is negatively charged. These charges exert a force on ions and molecules to keep them inside or outside of the cell. Just as ions and molecules diffuse down their concentration gradient the force of the electrical gradient (charge difference) between the extracellular and intracellular fluid affects the attraction or repulsion of the electrically charged ions and molecules on the opposite side of the membrane.

The Electrochemical Gradient

We have discussed simple concentration gradients and simple electrical gradients separately. However, in living systems gradients are more complex. Both gradients affect the movement of ions and molecules into and out of cells. These two gradients can work together or work against each other. The sum of these two forces, called the **electrochemical gradient**, determines the direction and speed of ion movement across the membrane ([link]).

Art Connection

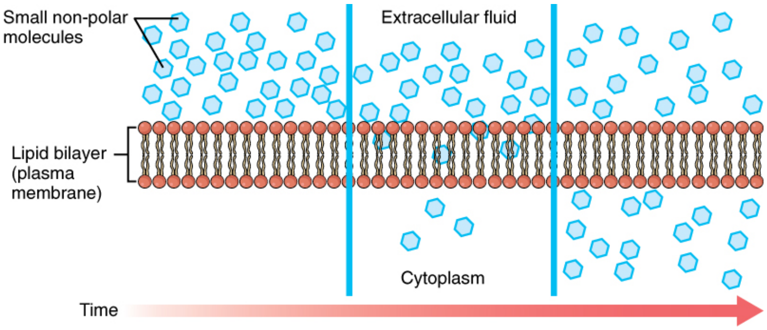


Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. This diagram shows these opposing forces on potassium ions inside the cell. (credit: Modified from image by “Synaptitude”/Wikimedia Commons)

When both forces work in the same direction their combined force makes for rapid movement of ions across the membrane. For example, there are more sodium (Na^+) ions outside the cell and the intracellular fluid is negatively charged so both forces work to "pull" the positively charged sodium ions into the cell. When the two forces oppose each other the movement is slowed or stopped. For example, there are more potassium (K^+) ions inside the cell so the forces of the concentration gradient "pull" the potassium out of the cell while the forces of the electrical gradient between the positively charged potassium ions and positively charged extracellular fluid "push" the potassium ions back inside the cell. The ratio of pull to push (electrochemical gradient) determines the effective force on the ions.

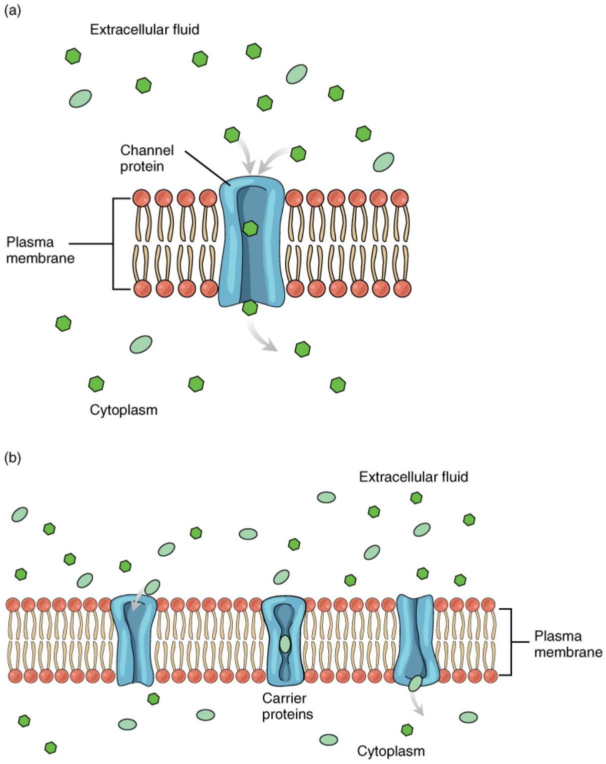
Passive Transport

Whenever a substance exists in greater concentration on one side of a semipermeable membrane, such as the cell membranes, any substance that can move down its concentration gradient across the membrane will do so spontaneously (without the addition of energy). Consider substances that can easily diffuse through the lipid bilayer of the cell membrane, such as the gases oxygen (O_2) and CO_2. Because cells rapidly use up oxygen during metabolism, there is typically a lower concentration of O_2 inside the cell than outside. As a result, oxygen will diffuse from the interstitial fluid directly through the lipid bilayer of the membrane and into the cytoplasm within the cell. On the other hand, because cells produce CO_2 as a byproduct of metabolism, CO_2 concentrations rise within the cytoplasm; therefore, CO_2 will move from the cell through the lipid bilayer and into the interstitial fluid, where its concentration is lower. This mechanism of molecules spreading from where they are more concentrated to where they are less concentration is a form of passive transport called **Simple diffusion** ([link]).



The structure of the lipid bilayer allows only small, non-polar substances such as oxygen and carbon dioxide to pass through the cell membrane, down their concentration gradient, by simple diffusion.

Solutes dissolved in water on either side of the cell membrane also will tend to diffuse down their concentration gradients, but because most substances cannot pass freely through the lipid bilayer of the cell membrane, their movement is restricted to protein channels and specialized transport mechanisms in the membrane. **Facilitated diffusion** is the diffusion process used for those substances that cannot cross the lipid bilayer due to their size and/or polarity ([link]). A common example of facilitated diffusion is the movement of glucose into the cell, where it is used to make ATP. Although glucose can be more concentrated outside of a cell, it cannot cross the lipid bilayer via simple diffusion because it is both large and polar. To resolve this, a specialized carrier protein called the glucose transporter will transfer glucose molecules into the cell to facilitate its inward diffusion.



(a) **Facilitated diffusion** of substances crossing the cell (plasma) membrane takes place with the help of proteins such as channel proteins and carrier proteins. **Channel proteins** are less selective than carrier proteins, and usually mildly discriminate between their cargo based on size and charge. (b) **Carrier proteins** are more selective, often only allowing one particular type of molecule to cross.

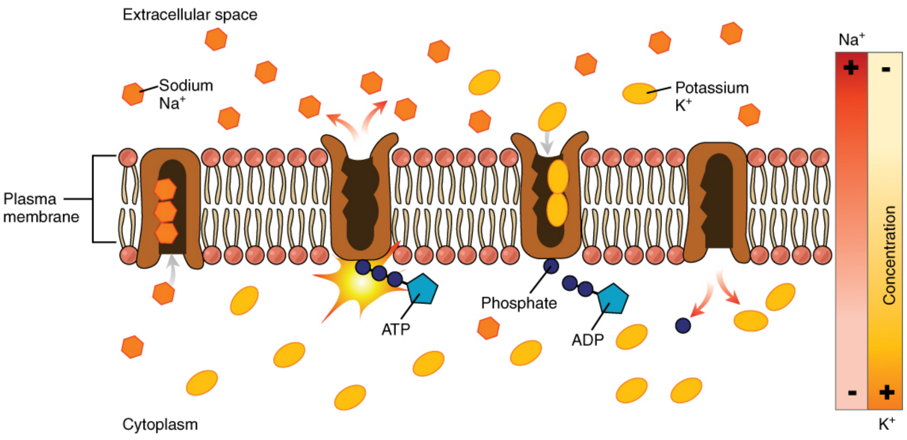
Another important example is sodium. Even though sodium ions (Na^+) are highly concentrated outside of cells, these electrolytes are polarized and cannot pass through the nonpolar lipid bilayer of the membrane. Their diffusion is facilitated by membrane proteins that form sodium channels (or "pores"), so that Na^+ ions can move down their concentration gradient from outside the cells to inside the cells. There are many other solutes that must undergo facilitated diffusion to move into a cell, such as amino acids, or to move out of a cell, such as wastes. Because facilitated diffusion is a passive process, it does not require energy expenditure by the cell.

Active Transport

To move substances against an electrochemical gradient, the cell must use energy. This energy generally is harvested from ATP generated through the cell's metabolism. Active transport mechanisms, collectively called **pumps**, work against electrochemical gradients. Small substances passively pass through plasma membranes all the time. Active transport maintains the proper concentrations of ions and other substances needed by living cells in the face of these passive movements. Much of a cell's supply of metabolic energy may be spent maintaining these processes. Most of a red blood cell's metabolic energy is used to maintain the imbalance between exterior and interior sodium and potassium levels required by the cell. Because active transport mechanisms depend on a cell's metabolism for energy, they are sensitive to many metabolic poisons that interfere with the supply of ATP.

Two mechanisms exist for the transport of small-molecular weight material and small molecules. **Primary active transport** uses ATP or some other high energy source to move ions or molecules across a membrane against their electrochemical gradient thereby maintaining or increasing the electrochemical gradient across that membrane. **Secondary active transport** uses the potential energy in the electrochemical gradient of one kind of ion or molecule established by the primary active transport to move a different kind of ion or molecule against its electrochemical gradient and does not use ATP directly.

Primary Active Transport uses a high energy molecule, like ATP, to power the pump. A common primary active transporter is the **sodium-potassium pump**, which is also called Na^+/K^+ ATPase. This pump transports sodium out of a cell while moving potassium into the cell. The Na^+/K^+ pump is an important ion pump found in the membranes of many types of cells. These pumps are particularly abundant in nerve cells, which are constantly pumping out sodium ions and pulling in potassium ions to maintain an electrical gradient across their cell membranes. In the case of nerve cells, for example, an electrical gradient exists between the inside and outside of the cell, with the inside being negatively-charged (at around -70 mV) relative to the outside. The negative electrical gradient is maintained because each Na^+/K^+ pump moves three Na^+ ions out of the cell and two K^+ ions into the cell for each ATP molecule that is used ([link]). This process is so important for nerve cells that it accounts for the majority of their ATP usage.



The sodium-potassium pump is found in many cell (plasma) membranes. Powered by ATP, the pump moves sodium and potassium ions in opposite directions, each against its concentration gradient. In a single cycle of the pump, three sodium ions are extruded from and two potassium ions are imported into the cell.

The Na^+-K^+ ATPase exists in two forms, depending on its orientation to the interior or exterior of the cell and its affinity for either sodium or potassium ions. The process consists of the following six steps.

- With the enzyme oriented towards the interior of the cell, the carrier has a high affinity for sodium ions. Three ions bind to the protein.
- ATP is hydrolyzed by the protein carrier and a low-energy phosphate group attaches to it.
- As a result, the carrier changes shape and re-orients itself towards the exterior of the membrane. The protein’s affinity for sodium decreases and the three sodium ions leave the carrier.
- The shape change increases the carrier’s affinity for potassium ions, and two such ions attach to the protein. Subsequently, the low-energy phosphate group detaches from the carrier.
- With the phosphate group removed and potassium ions attached, the carrier protein repositions itself towards the interior of the cell.
- The carrier protein, in its new configuration, has a decreased affinity for potassium, and the two ions are released into the cytoplasm. The protein now has a higher affinity for sodium ions, and the process starts again.

Several things have happened as a result of this process. At this point, there are more sodium ions outside of the cell than inside and more potassium ions inside than out. For every three ions of sodium that move out, two ions of potassium move in. This results in the interior being slightly more negative relative to the exterior. This difference in charge is important in creating the conditions necessary for the secondary process. The sodium-potassium pump is, therefore, an **electrogenic pump** (a pump that creates a charge imbalance), creating an electrical imbalance across the membrane and contributing to the membrane potential.

Note:
Link to Learning

Visit the site to see a simulation of active transport in a sodium-potassium ATPase.

Note:
Cell: Cystic Fibrosis
Cystic fibrosis (CF) affects approximately 30,000 people in the United States, with about 1,000 new cases reported each year. The genetic disease is most well known for its damage to the lungs, causing breathing difficulties and chronic lung infections, but it also affects the liver, pancreas, and intestines. Only about 50 years ago, the prognosis for children born with CF was very grim—a life expectancy rarely over 10 years. Today, with advances in medical treatment, many CF patients live into their 30s.

The symptoms of CF result from a malfunctioning membrane ion channel called the cystic fibrosis transmembrane conductance regulator, or CFTR. In healthy people, the CFTR protein is an integral membrane protein that transports Cl^- ions out of the cell. In a person who has CF, the gene for the CFTR is mutated, thus, the cell manufactures a defective channel protein that typically is not incorporated into the membrane, but is instead degraded by the cell.

The CFTR requires ATP in order to function, making its Cl^- transport a form of active transport. This characteristic puzzled researchers for a long time because the Cl^- ions are actually flowing *down* their concentration gradient when transported out of cells. Active transport generally pumps ions *against* their concentration gradient, but the CFTR presents an exception to this rule.

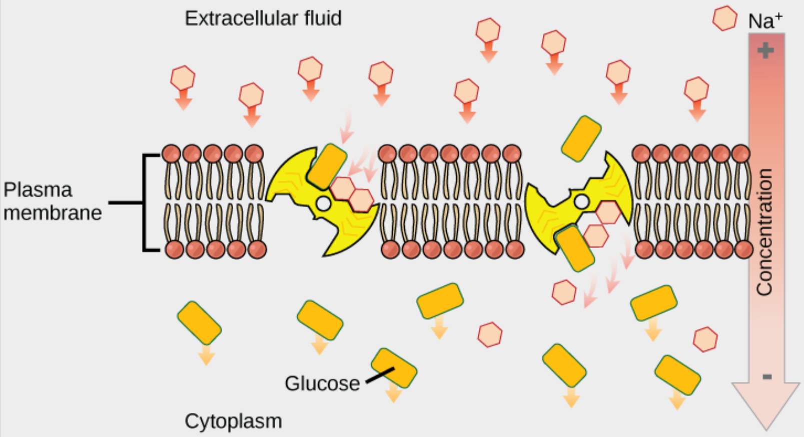
In normal lung tissue, the movement of Cl^- out of the cell maintains a Cl^--rich, negatively charged environment immediately outside of the cell. This is particularly important in the epithelial lining of the respiratory system. Respiratory epithelial cells secrete mucus, which serves to trap dust, bacteria, and other debris. A cilium (plural = cilia) is one of the hair-like appendages found on certain cells. Cilia on the epithelial cells move the mucus and its trapped particles up the airways away from the lungs and toward the outside. In order to be effectively moved upward, the mucus cannot be too viscous; rather it must have a thin, watery consistency. The transport of Cl^- and the maintenance of an electronegative environment outside of the cell attract positive ions such as Na^+ to the extracellular space. The accumulation of both Cl^- and Na^+ ions in the extracellular space creates solute-rich mucus, which has a low concentration of water molecules. As a result, through osmosis, water moves from cells and extracellular matrix into the mucus, “thinning” it out. This is how, in a normal respiratory system, the mucus is kept sufficiently watered-down to be propelled out of the respiratory system.

If the CFTR channel is absent, Cl^- ions are not transported out of the cell in adequate numbers, thus preventing them from drawing positive ions. The absence of ions in the secreted mucus results in the lack of a normal water concentration gradient. Thus, there is no osmotic pressure pulling water into the mucus. The resulting mucus is thick and sticky, and the ciliated epithelia cannot effectively remove it from the respiratory system. Passageways in the lungs become blocked with mucus, along with the debris it carries. Bacterial infections occur more easily because bacterial cells are not effectively carried away from the lungs.

Active transport pumps can also work together with other active or passive transport systems to move substances across the membrane. For example, the sodium-potassium pump maintains a high concentration of sodium ions outside of the cell. Therefore, if the cell needs sodium ions, all it has to do is open a passive sodium channel, as the concentration gradient of the sodium ions will drive them to diffuse into the cell. In this way, the action of an active transport pump (the sodium-potassium pump) powers the passive transport of sodium ions by creating a concentration gradient. When active transport powers the transport of a different substance in this way, it is called secondary active transport.

Secondary Active Transport uses the diffusion energy of sodium ions or other molecules to pump a different molecule or ion across the plasma membrane against its concentration gradient. As sodium ion concentrations build outside of the plasma membrane because of the action of the primary active transport process, an electrochemical gradient is created. If a channel protein exists and is open, the sodium ions will be pulled through the membrane. This movement is used to transport other substances through the membrane against their electrochemical gradient by attaching them to the same transport protein ([link]). Many amino acids, as well as glucose, enter a cell this way. This secondary process is also used to store high-energy hydrogen ions in the mitochondria of plant and animal cells for the production of ATP. The potential energy that accumulates in the stored hydrogen ions is translated into kinetic energy as the ions surge through the channel protein ATP synthase, and that energy is used to convert ADP into ATP.

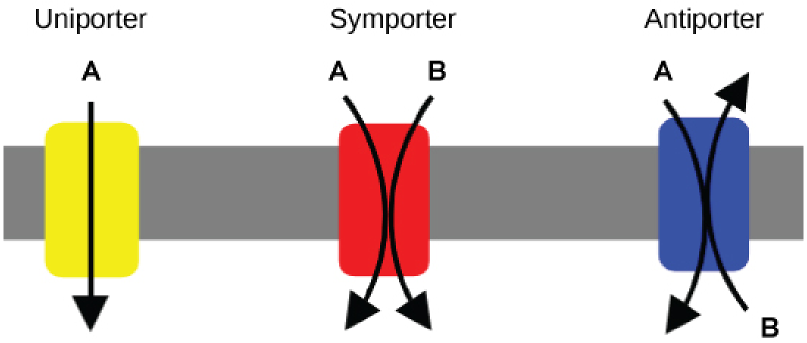
Note:
Art Connection



The conversion of potential energy in an electrochemical gradient, *created by primary active transport*, to diffusion energy can move other substances against their concentration gradients, a process called *secondary active transport*. (credit: modification of work by Mariana Ruiz Villareal)

If the pH outside the cell decreases, would you expect the amount of amino acids transported into the cell to increase or decrease?

An important membrane adaption for active transport is the presence of specific carrier proteins or pumps to facilitate movement: there are three types of these proteins or **transporters** ([link]). A **uniporter** carries one specific ion or molecule. A **symporter** carries two different ions or molecules, both in the same direction. An **antiporter** also carries two different ions or molecules, but in different directions. All of these transporters can also transport small, uncharged organic molecules like glucose. These three types of carrier proteins are also found in facilitated diffusion, but they do not require ATP to work in that process. Some examples of pumps for active transport are Na^+-K^+ ATPase, which carries sodium and potassium ions, and H^+-K^+ ATPase, which carries hydrogen and potassium ions. Both of these are antiporter carrier proteins. Two other carrier proteins are Ca^{2+} ATPase and H^+ ATPase, which carry only calcium and only hydrogen ions, respectively. Both are pumps.



A uniporter carries one molecule or ion. A symporter carries two different molecules or ions, both in the same direction. An antiporter also carries two different molecules or ions, but in different directions. (credit: modification of work by “Lupask”/Wikimedia Commons)

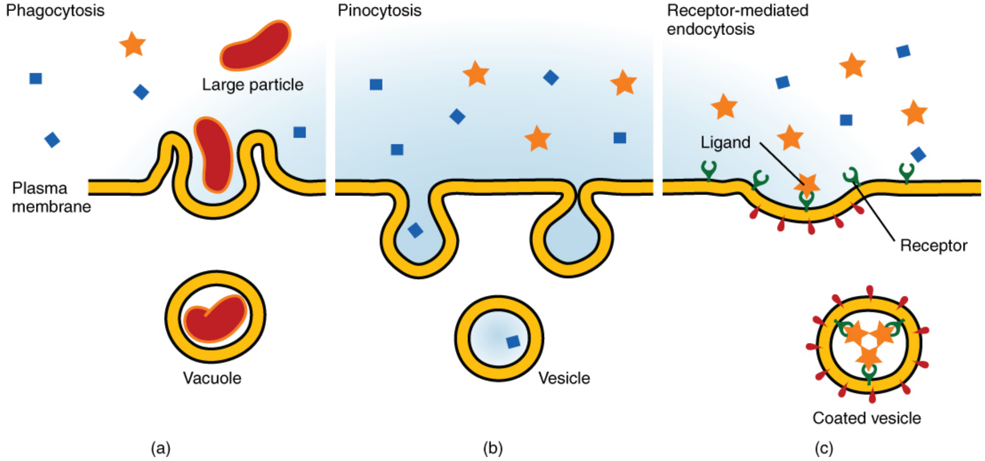
Symporters are secondary active transporters that move two substances in the same direction. For example, the sodium-glucose symporter uses sodium ions to “pull” glucose molecules into the cell. Because cells store glucose for energy, glucose is typically at a higher concentration inside of the cell than outside. However, due to the action of the sodium-potassium pump, sodium ions will easily diffuse into the cell when the symporter is opened. The flood of sodium ions through the symporter provides the energy that allows glucose to move through the symporter and into the cell, against its concentration gradient.

Conversely, antiporters are secondary active transport systems that transport substances in opposite directions. For example, the sodium-hydrogen ion antiporter uses the energy from the inward flood of sodium ions to move hydrogen ions (H+) out of the cell. The sodium-hydrogen antiporter is used to maintain the pH of the cell's interior.

Bulk Transport

Other forms of active transport do not involve membrane carriers. **Endocytosis** (bringing “into the cell”) is the process of a cell ingesting material by enveloping it in a portion of its cell membrane, and then pinching off that portion of membrane ([link]). Once pinched off, the portion of membrane and its contents becomes an independent, intracellular vesicle. A **vesicle** is a membranous sac—a spherical and hollow organelle bounded by a lipid bilayer membrane. Endocytosis often brings materials into the cell that must to be broken down or digested. **Phagocytosis** (“cell eating”) is the endocytosis of large particles. Many immune cells engage in phagocytosis of invading pathogens. Like little Pac-men, their job is to patrol body tissues for unwanted matter, such as invading bacterial cells, phagocytize them, and digest them. In contrast to phagocytosis, **pinocytosis** (“cell drinking”) brings fluid containing dissolved substances into a cell through membrane vesicles.

Three Forms of Endocytosis

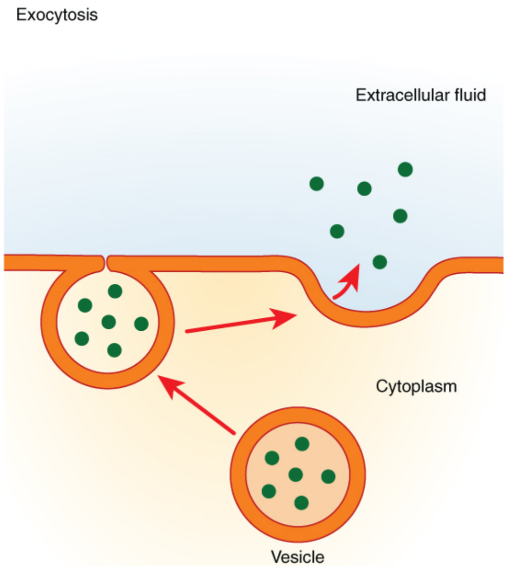


Endocytosis is a form of active transport in which a cell envelopes extracellular materials using its cell membrane. (a) In phagocytosis, which is relatively nonselective, the cell takes in a large particle. (b) In pinocytosis, the cell takes in small particles in fluid. (c) In contrast, receptor-mediated endocytosis is quite selective. When external receptors bind a specific ligand, the cell responds by endocytosing the ligand.

Phagocytosis and pinocytosis take in large portions of extracellular material, and they are typically not highly selective in the substances they bring in. Cells regulate the endocytosis of specific substances via receptor-mediated endocytosis. **Receptor-mediated endocytosis** is endocytosis by a portion of the cell membrane that contains many receptors that are specific for a certain substance. Once the surface receptors have bound sufficient amounts of the specific substance (the receptor's ligand), the cell will endocytose the part of the cell membrane containing the receptor-ligand complexes. Iron, a required component of hemoglobin, is endocytosed by red blood cells in this way. Iron is bound to a protein called transferrin in the blood. Specific transferrin receptors on red blood cell surfaces bind the iron-transferrin molecules, and the cell endocytoses the receptor-ligand complexes.

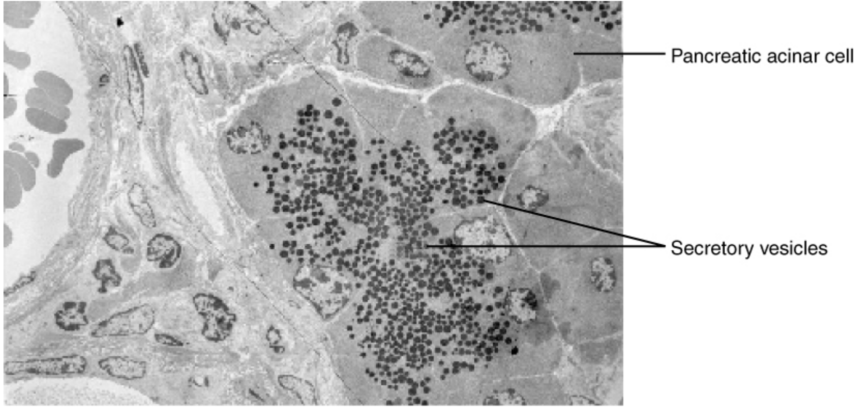
In contrast with endocytosis, **exocytosis** (taking "out of the cell") is the process of a cell exporting material using vesicular transport ([link]). Many cells manufacture substances that must be secreted, like a factory manufacturing a product for export. These substances are typically packaged into membrane-bound vesicles within the cell. When the vesicle membrane fuses with the cell membrane, the vesicle releases it contents into the interstitial fluid. The vesicle membrane then becomes part of the cell membrane. Cells of the stomach and pancreas produce and secrete digestive enzymes through exocytosis ([link]). Endocrine cells produce and secrete hormones that are sent throughout the body, and certain immune cells produce and secrete large amounts of histamine, a chemical important for immune responses.

Exocytosis



Exocytosis is much like endocytosis in reverse. Material destined for export is packaged into a vesicle inside the cell. The membrane of the vesicle fuses with the cell membrane, and the contents are released into the extracellular space.

Pancreatic Cells' Enzyme Products



The pancreatic acinar cells produce and secrete many enzymes that digest food. The tiny black granules in this electron micrograph are secretory vesicles filled with enzymes that will be exported from the cells via exocytosis. LM × 2900. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

Note:

View the University of Michigan WebScope at http://virtualslides.med.umich.edu/Histology/EMsmallCharts/3%20Image%20Scope%20finals/226%20-%20Pancreas_001.svs/view.apml to explore the tissue sample in greater detail.

Moving Water: Osmosis

Up to this point we have been looking at the movement of ions and molecules that are dissolved or in suspension. But the fluid through which they are moving also is made up of molecules. While all the molecules involved follow the same physical laws, the concentration of molecules of the solvent (the fluid in which the ions and molecules are dissolved) in a given area is the inverse of the concentration of solute in that area. Put more simply, no two molecules can occupy the exact same space at the same time. Therefore, the more solute molecules in a given space, the fewer solvent molecules there can be in the same space, and vice versa. In biological systems the solvent is water, so the following discussion will use water rather than the more general term, solvent.

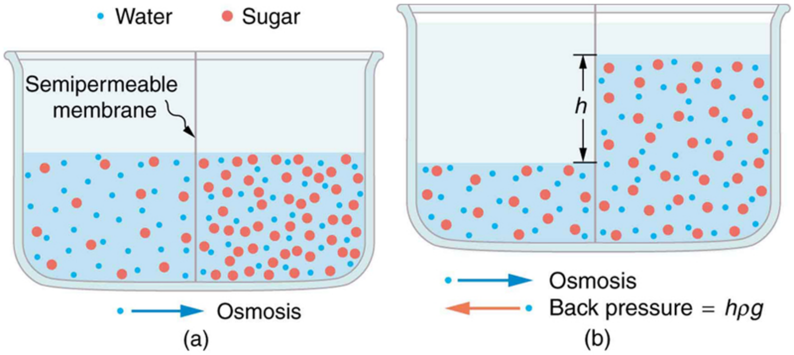
Just as the solute molecules diffuse from areas where they are in a higher concentration to areas of lower concentration so too do the water molecules move from their areas of higher concentration to areas of their lower concentration. However, the direction of the molecular diffusion for the water is opposite that of the solute. Remember the area with a high concentration of solute molecules will necessarily have a lower concentration of water molecules.

While the underlying driving forces are the same for both solute and water, in simple diffusion we tend to concentrate on the movement of the solute and ignore the water molecules diffusing in the opposite direction. However, if we use a semipermeable membrane (one that lets the water molecules diffuse freely while the solute molecules are restricted to their side) to separate two solutions with differing concentrations of solute, we can see and measure the independent diffusion of the water molecules down their concentration gradient.

Osmosis is the movement of water through a semipermeable membrane from a region of high concentration to a region of low concentration. Osmosis is driven by the imbalance in water concentration. For example, water is more concentrated in your body than in Epsom salt. When you soak a swollen ankle in Epsom salt, the water moves out of your body into the lower-concentration region in the salt. In contrast, **dialysis** is the transport of any other molecule through a semipermeable membrane due to its concentration difference. Both osmosis and dialysis are used by the kidneys to cleanse the blood.

Osmosis can create a substantial pressure. Consider what happens if osmosis continues for some time, as illustrated in [link]. Water moves by osmosis from the left into the region on the right, where it is less concentrated, causing the solution on the right to rise. This movement will continue until the pressure ρgh created by the weight of the extra fluid on the right is large enough to stop further osmosis. This pressure is called a *back pressure*. The back pressure ρgh that stops osmosis is also called

osmotic pressure. Osmotic pressure can be large, depending on the size of the concentration difference. For example, if pure water and sea water are separated by a semipermeable membrane that passes no salt, osmotic pressure will be 25.9 atm. This value means that water will diffuse through the membrane until the salt water surface rises 268 meters above the pure-water surface! One example of pressure created by osmosis is turgor in plants (many wilt when too dry). Turgor describes the condition of a plant in which the fluid in a cell exerts a pressure against the cell wall. This pressure gives the plant support. Dialysis can similarly cause substantial pressures.



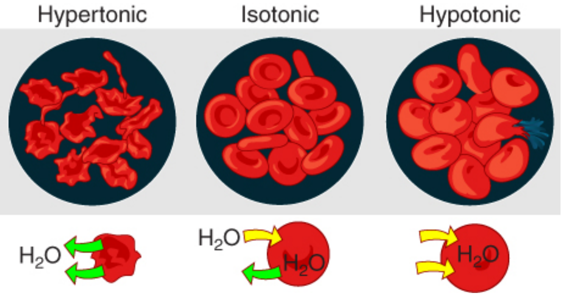
(a) Two sugar-water solutions of different concentrations, separated by a semipermeable membrane that passes water but not sugar. Osmosis will be to the right, since water is less concentrated there. (b) The fluid level rises until the back pressure ρgh equals the relative osmotic pressure; then, the net transfer of water is zero.

In physiology we generally measure the osmotic pressure in terms of osmoles (Osm) rather than atmospheres (atm). An osmole is a unit of measurement that defines the number of moles of solute that contribute to the osmotic pressure of a solution. For example, a 1 molar solution (1mol/L) of NaCl would create a 2 osmole solution because the sodium and chloride in solution separate into two separate particles and it is the number of particles that create the osmotic gradient. Glucose molecules by comparison do not separate in solution so a 1 molar solution of glucose would yield a 1 osmole solution.

The movement of water molecules is not itself regulated by cells, so it is important that cells are exposed to an environment in which the concentration of solutes outside of the cells (in the extracellular fluid) is equal to the concentration of solutes inside the cells (in the cytoplasm). Two solutions that have the same concentration of solutes are said to be **isotonic** (equal tension). When cells and their extracellular environments are isotonic, the concentration of water molecules is the same outside and inside the cells, and the cells maintain their normal shape (and function). In most of the human body this is a concentration around 300 mOsm (milliosmoles) or 0.3 osmoles.

Osmosis occurs when there is an imbalance of solutes outside of a cell versus inside the cell. A solution that has a higher concentration of solutes than another solution is said to be **hypertonic**, and water molecules tend to diffuse into a hypertonic solution ([link]). Cells in a hypertonic solution will shrivel as water leaves the cell via osmosis. In contrast, a solution that has a lower concentration of solutes than another solution is said to be **hypotonic**, and water molecules tend to diffuse out of a hypotonic solution and into the cells. Cells in a hypotonic solution will take on too much water and swell, with the risk of eventually bursting. A critical aspect of homeostasis in living things is to create an internal environment in which all of the body's cells are in an isotonic solution. Various organ systems, particularly the kidneys, work to maintain this homeostasis.

Concentration of Solutions



A hypertonic solution has a solute concentration higher than another solution. An isotonic solution has a solute concentration equal to another solution. A hypotonic solution has a solute concentration lower than another solution.

Another mechanism besides diffusion to passively transport materials between compartments is filtration. Unlike diffusion of a substance from where it is more concentrated to less concentrated, filtration uses a hydrostatic pressure (water pressure) gradient that pushes the fluid and the solutes within it from a higher pressure area to a lower pressure area. Filtration is an extremely important process in the body. For example, the circulatory system uses filtration to move plasma and substances across the endothelial lining of capillaries and into surrounding tissues, supplying cells with the nutrients. Filtration pressure in the kidneys provides the mechanism to remove wastes from the bloodstream.

In humans water is moved passively either by osmosis or hydrostatic pressure. However, other organisms do move water using active transport mechanisms. Cypress tree roots, for example, extract pure water from salt water, although osmosis would move it in the opposite direction. This is not reverse osmosis, because there is no back pressure to cause it.

Thought Experiment

Exercise:

Problem:

Remember everyone started standing close together in one corner of the room. When the lights are turned on do you think they will all be in the same corner of the room? If not, how do you think they will be distributed around the room? Why?

Solution:

After fifteen minutes everyone will be distributed evenly around the room, even though they cannot see where they are or the location of anyone else. This will happen because as each person bumps into another person the tendency is to move in the opposite direction. Because each person is more likely to bump into another person in the areas of highest concentration (where there are more people to bump into) and less likely to bump into another person in areas of lower concentration, they will move further in directions of lower concentration before they bump into another person. As everyone is doing this, eventually each individual will wind up with roughly equal distance in every direction between him or her self and the individuals around them.

Exercise:

Problem:

Do you think the speed at which everyone is moving (walking slow versus walking fast) would have an effect on how quickly they reach equal distribution? The speed of movement of molecules is directly related to the temperature of the solution. What conclusions can you draw about temperature and diffusion?

Solution:

If everyone is moving faster, they will bump into each other more often, so the even distribution would occur more quickly. This is why diffusion happens more rapidly as temperatures increase and more slowly as they decrease.

Exercise:

Problem:

Would there be a difference in how quickly equal distribution would be reached if we ran the same experiment twice, but started with twice as many people in the corner in the second trial? What conclusions can you draw about concentration (number of people) in the corner and diffusion?

Solution:

With more people packed into the same space in the second trial, they will bump into each other more often than they did in the first trial. This increase in bumping will increase the rate at which they distribute around the room. This is why the greater the *Concentration Gradient* (difference in solute concentration between areas of the solution) the faster the diffusion will happen.

Art Connections

Exercise:

Problem:

If the pH outside the cell decreases, would you expect the amount of amino acids transported into the cell to increase or decrease?

Solution:

A decrease in pH means an increase in positively charged H^+ ions, and an increase in the electrical gradient across the membrane. The transport of amino acids into the cell will increase.

Chapter Review

The cell membrane provides a barrier around the cell, separating its internal components from the extracellular environment. It is composed of a phospholipid bilayer, with hydrophobic internal lipid “tails” and hydrophilic external phosphate “heads.” Various membrane proteins are scattered throughout the bilayer, both inserted within it and attached to it peripherally. The cell membrane is selectively permeable, allowing only a limited number of materials to diffuse through its lipid bilayer. All materials that cross the membrane do so using passive (non energy-requiring) or active (energy-requiring) transport processes. During passive transport, materials move by simple diffusion or by facilitated diffusion through the membrane, down their concentration gradient. Water passes through the membrane in a diffusion process called osmosis. During active transport, energy is expended to assist material movement across the membrane in a direction against their concentration gradient. Active transport may take place with the help of protein pumps or through the use of vesicles.

Interactive Link Questions

Exercise:

Problem:

Visit this link to see diffusion and how it is propelled by the kinetic energy of molecules in solution. How does temperature affect diffusion rate, and why?

Solution:

Higher temperatures speed up diffusion because molecules have more kinetic energy at higher temperatures.

Review Questions

Exercise:

Problem:

Because they are embedded within the membrane, ion channels are examples of _________.

a. receptor proteins
b. integral proteins
c. peripheral proteins
d. glycoproteins

Solution:

B

Exercise:

Problem:

The diffusion of substances within a solution tends to move those substances ________ their ________ gradient.

a. up; electrical
b. up; electrochemical
c. down; pressure
d. down; concentration

Solution:

D

Exercise:

Problem: Ion pumps and phagocytosis are both examples of ________.

a. endocytosis
b. passive transport
c. active transport
d. facilitated diffusion

Solution:

C

Exercise:

Problem:

Choose the answer that best completes the following analogy: Diffusion is to ________ as endocytosis is to ________.

a. filtration; phagocytosis

b. osmosis; pinocytosis
c. solutes; fluid
d. gradient; chemical energy

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

What materials can easily diffuse through the lipid bilayer, and why?

Solution:

Only materials that are relatively small and nonpolar can easily diffuse through the lipid bilayer. Large particles cannot fit in between the individual phospholipids that are packed together, and polar molecules are repelled by the hydrophobic/nonpolar lipids that line the inside of the bilayer.

Exercise:

Problem:

Why is receptor-mediated endocytosis said to be more selective than phagocytosis or pinocytosis?

Solution:

Receptor-mediated endocytosis is more selective because the substances that are brought into the cell are the specific ligands that could bind to the receptors being endocytosed. Phagocytosis or pinocytosis, on the other hand, have no such receptor-ligand specificity, and bring in whatever materials happen to be close to the membrane when it is enveloped.

Exercise:

Problem:

What do osmosis, diffusion, filtration, and the movement of ions away from like charge all have in common? In what way do they differ?

Solution:

These four phenomena are similar in the sense that they describe the movement of substances down a particular type of gradient. Osmosis and diffusion involve the movement of water and other substances down their concentration gradients, respectively. Filtration describes the movement of particles down a pressure gradient, and the movement of ions away from like charge describes their movement down their electrical gradient.

Glossary

active transport

form of transport across the cell membrane that requires input of cellular energy

amphipathic

describes a molecule that exhibits a difference in polarity between its two ends, resulting in a difference in water solubility

cell membrane

membrane surrounding all animal cells, composed of a lipid bilayer interspersed with various molecules; also known as plasma membrane

channel protein

membrane-spanning protein that has an inner pore which allows the passage of one or more substances

concentration gradient
difference in the concentration of a substance between two regions

diffusion
movement of a substance from an area of higher concentration to one of lower concentration

electrical gradient
difference in the electrical charge (potential) between two regions

endocytosis
import of material into the cell by formation of a membrane-bound vesicle

exocytosis
export of a substance out of a cell by formation of a membrane-bound vesicle

extracellular fluid (ECF)
fluid exterior to cells; includes the interstitial fluid, blood plasma, and fluid found in other reservoirs in the body

facilitated diffusion
diffusion of a substance with the aid of a membrane protein

glycocalyx
coating of sugar molecules that surrounds the cell membrane

glycoprotein
protein that has one or more carbohydrates attached

hydrophilic
describes a substance or structure attracted to water

hydrophobic
describes a substance or structure repelled by water

hypertonic

describes a solution concentration that is higher than a reference concentration

hypotonic
describes a solution concentration that is lower than a reference concentration

integral protein
membrane-associated protein that spans the entire width of the lipid bilayer

interstitial fluid (IF)
fluid in the small spaces between cells not contained within blood vessels

intracellular fluid (ICF)
fluid in the cytosol of cells

isotonic
describes a solution concentration that is the same as a reference concentration

ligand
molecule that binds with specificity to a specific receptor molecule

osmosis
diffusion of molecules down their concentration across a selectively permeable membrane

passive transport
form of transport across the cell membrane that does not require input of cellular energy

peripheral protein
membrane-associated protein that does not span the width of the lipid bilayer, but is attached peripherally to integral proteins, membrane lipids, or other components of the membrane

phagocytosis
endocytosis of large particles

pinocytosis
endocytosis of fluid

receptor
protein molecule that contains a binding site for another specific molecule (called a ligand)

receptor-mediated endocytosis
endocytosis of ligands attached to membrane-bound receptors

selective permeability
feature of any barrier that allows certain substances to cross but excludes others

sodium-potassium pump
(also, Na^+/K^+ ATP-ase) membrane-embedded protein pump that uses ATP to move Na^+ out of a cell and K^+ into the cell

vesicle
membrane-bound structure that contains materials within or outside of the cell

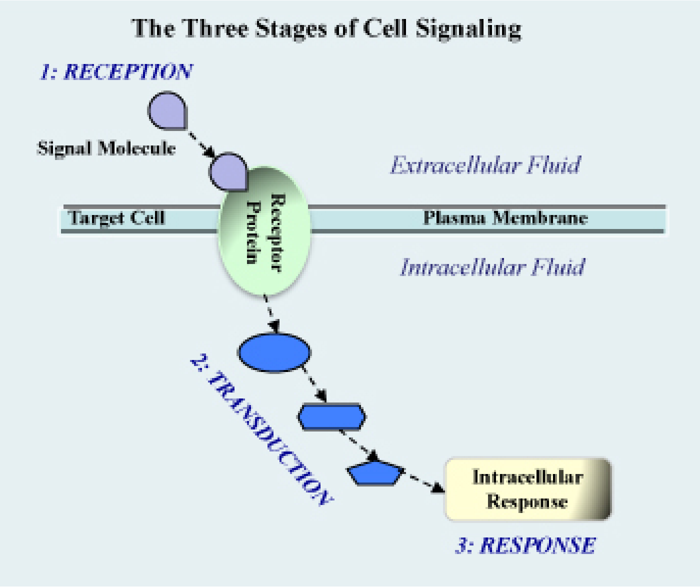
Introduction to Cellular Communication

Why Cells Communicate

Both internal and external cellular communication are essential for living organisms. Internally cells need to communicate between the various biochemical steps and processes carried out within the cell. Externally cells need to communicate with other cells to coordinate activities, whether they are bacterial colonies or differentiated cells coordinating infrastructure activities within the organism. Recent studies also have demonstrated important communication taking place between bacterial cells (often in the GI tract) and cells in other areas of their host organism. All of these forms of cellular communication use similar mechanisms. It is worth keeping in mind that communication refers to the Signal -> Target -> Response activities at the chemical level and does not imply any sort of conscious activity or understanding on the part of the participating cells or proteins.

At the highest level there are two kinds of communication in the world of living cells. Communication between cells is called **intercellular signaling**, and communication within a cell is called **intracellular signaling**. An easy way to remember the distinction is by understanding the Latin origin of the prefixes: inter- means "between" (for example, intersecting lines are those that cross each other) and intra- means "inside" (like intravenous). Often the intracellular communication processes initially are triggered by extracellular signals.

Biological communication has several steps. In intercellular communication the target cell must convert the *external signal* into an *internal response* in order for the communication to be successful. The originator of the communication generates a message called a **signal**. The signal must be received by the **target**. The target cell then **transduces** the signal into a **response**. ([link])



Communication between two cells involves three stages: Reception of the signal from the initiating cell, Transduction of the signal via intracellular pathways, and a Response (change) in the receiving cell.

The initial cellular response generally involves opening a channel or activating either an internal enzyme or a high-energy chemical molecule. Generating the response often requires additional internal communication, where the first response becomes a signal for an internal target protein that responds by creating one or more additional signals for other proteins. This intracellular communication process continues until the final cellular response is achieved. This is an example of the generic cause-and-effect-chain process discussed in the first chapter, *An Introduction to the Human Body*. In this context the cause and effect chain is called a **transduction pathway**.

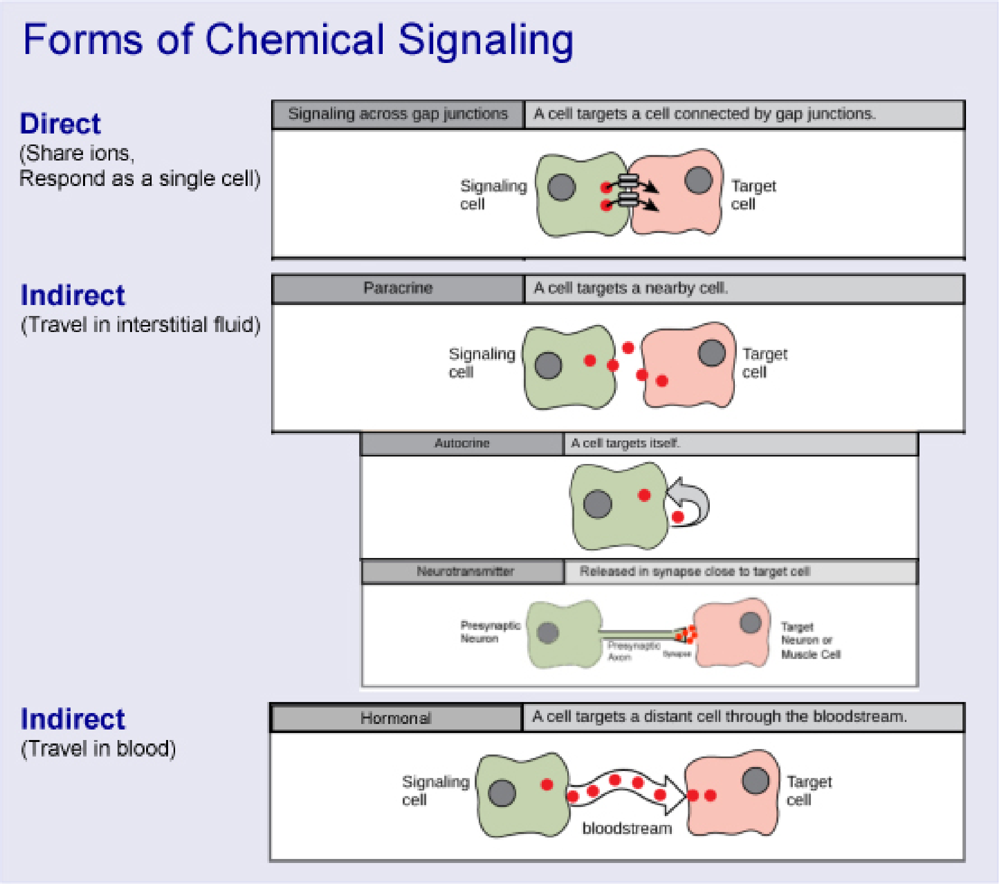
Chemical signals are released by **signaling cells** in the form of small, usually volatile or soluble molecules called ligands. A **ligand** is a molecule that binds with another specific molecule, in some cases, delivering a signal in the process. Ligands can thus be thought of as signaling molecules. Ligands interact with proteins, called **receptors**, in **target cells**, which are cells that are affected by the chemical signal. Ligands and receptors exist in several varieties; however, a specific ligand will have a specific receptor that typically binds only with that ligand or chemicals similar to that ligand (for example agonists and antagonists). You may want to review the section on enzymes in the *Cellular Metabolism* chapter as the ligand/receptor part of the process works in a similar way in both activities.

Intercellular Communication Mechanisms

There are two fundamentally different forms of communication between cells, **direct and indirect**. Direct communication occurs between cells that are physically connected. The signals pass from one cell to another through cytoplasmic fluid. Indirect communication involves chemical diffusion across the interstitial fluid between cells. ([link])

Direct communication in animal cells takes place through gap junctions. These water-filled channels between the cells allow small signaling molecules, called **intracellular mediators**, to diffuse between the two cells. Small molecules, such as calcium ions (Ca^{2+}), are able to move between cells, but large molecules like proteins and DNA cannot fit through the channels. The specificity of the channels ensures that the cells remain independent but can quickly and easily transmit signals. The transfer of signaling molecules communicates the current state of the cell that is directly next to the target cell; this allows a group of cells to coordinate their response to a signal that only one of them may have received. In plants, the equivalent of gap junctions (plasmodesmata) are ubiquitous, making the entire plant into a giant, communication network.

Indirect communication always involves the release of the signal into interstitial fluid outside the signaling cell and diffusion of the signal molecules to surrounding cells. With one major exception the target cells are within diffusion distance of the signaling cell. The exception is found in signal molecules that are released into an area of dense capillary beds. In this case the signal molecules diffuse from the interstitial fluid into the blood stream where they are distributed into the interstitial fluid throughout the body.


Forms of Chemical Signaling
Direct
(Share ions,
Respond as a single cell)
Signaling across gap junctions
A cell targets a cell connected by gap junctions.
Signaling cell
Target cell
Indirect
(Travel in interstitial fluid)
Paracrine
A cell targets a nearby cell.
Signaling cell
Target cell
Autocrine
A cell targets itself.
Neurotransmitter
Released in synapse close to target cell
Presynaptic Neuron
Presynaptic Axon
Synapse
Target Neuron or Muscle Cell
Indirect
(Travel in blood)
Hormonal
A cell targets a distant cell through the bloodstream.
Signaling cell
bloodstream
Target cell

Cell signaling can be organized into a hierarchy of categories: **Direct** vs. **Indirect** > **Indirect** = Paracrine vs. Hormonal > **Paracrine** = General vs. Autocrine vs. Neuronal.

Hormonal Signaling

Hormones are ligands that reach their targets by traveling in the blood. As such, they are signaling molecules that are produced in one part of the body but affect cells in other body regions some distance away. In the body, many hormones originate from **endocrine cells**. Endocrine cells are located in endocrine glands, such as the thyroid gland, the hypothalamus, and the pituitary gland. However, cells in non-endocrine organs, such as the heart, kidney, and adipose tissue, also release hormones into the blood.

Hormones travel long distances between the signaling cells that released them and their target cells via the bloodstream, which is a relatively slow way to move throughout the body. While these blood-borne signals usually produce a slower response they also have a longer-lasting effect. The same hormone also can affect multiple organs with one release. Because of their form of transport, hormones get diluted and are present in low concentrations when they act on their target cells. This is different from paracrine signaling, in which local concentrations of ligands can be very high.

Paracrine Signaling

Signals that act locally between cells that are close together are called **paracrine signals**. Paracrine signals move by diffusion through the extracellular matrix. Although paracrines may enter the bloodstream, their concentration is generally too low to elicit a response from distant tissues.

A familiar paracrine example experienced by those with allergies is histamine, a paracrine released by immune cells (basophils and mast cells). Histamines increase the permeability of the capillaries to white blood cells and some proteins allowing them to engage pathogens in infected tissues. When the immune system mistakenly targets non-pathogenic material the histamine release causes the symptoms (excess fluid, runny nose, and watery eyes) seen in allergies. Antihistamines are taken to counter the effects of this paracrine.

Paracrine communication also is critical for cell differentiation during development. The paracrine signals released by surrounding cells determine the specific differentiation pathway of developing stem cells. These same paracrines either released by mature cells in the area of placement or artificially introduced guide the correct differentiation and development of embryonic stem cells in stem cell therapies.

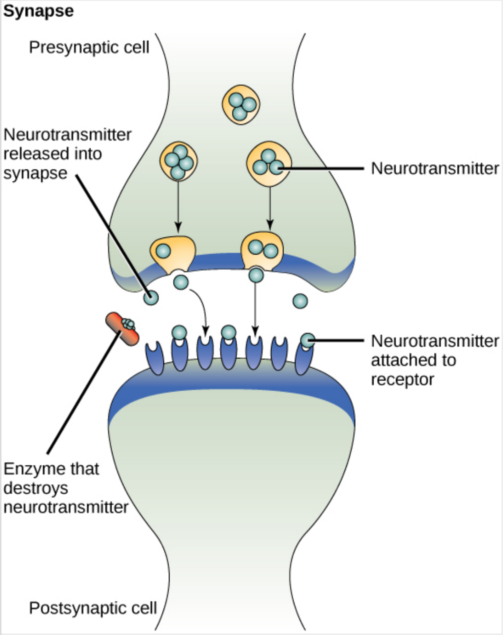
In addition to general paracrine communication with the local cells there are two special types of paracrine action: autocrine and neuronal.

Autocrine communication occurs when the cell releasing the paracrine also has receptors for and responds to the paracrine it is releasing. This is a form of self communication, but instead of communicating using intracellular pathways the cell "talks to itself" via extracellular mechanisms. The prefix *auto-* means self, a reminder that the signaling cell sends a signal to itself. This type of signaling often occurs during the early development of an organism to ensure that cells develop into the correct tissues and take on the proper function. In some cases, neighboring cells of the same type are also influenced by the released ligand. In embryological development, this process of stimulating a group of neighboring cells may help to direct the differentiation of identical cells into the same cell type, thus ensuring the proper developmental outcome. In the immune response autocrine signals, called interleukins, create a positive feedback loop in T-Helper cells rapidly amplifying the response. Autocrine signaling also regulates pain sensation and inflammatory responses. Further, if a cell is infected with a virus, the cell can signal itself to undergo programmed cell death, killing the virus in the process.

Neuronal communication allows the signaling cells (neurons) to release the chemical ligands, called **neurotransmitters**, a long distance from the cell body. Physical extension of the cell itself (the axon) spans the long distance to a specific target so the ligand is released in a space very close to the target cell called a **synapse**. The ability to communicate with distant targets is a trait shared with hormonal communication. However, unlike the widespread distribution of hormonal ligands the release of the neurotransmitter happens very close to the target cell and only affects a small area near the target. This makes the concentration of neurotransmitter near the target receptors much higher than the concentration of ligand around hormonal target receptors. The ligand itself only diffuses a short distance to the target and is rapidly broken down or removed from the area of release, making it a form of paracrine communication.

Note:

A nerve cell consists of a cell body, several short, branched extensions called dendrites that receive stimuli, and a long extension called an axon, which transmits signals to other nerve cells or muscle cells. The junction between nerve cells where signal transmission occurs is called a **synapse**. Signals within the nerve cells are propagated down the axon by fast-moving electrical impulses. When these impulses reach the end of the presynaptic cell's axon, the signal is communicated to the postsynaptic cell by the release of chemical ligands called **neurotransmitters**. The neurotransmitters diffuse across the very small distances between nerve cells (synapses) ([link]). The small distance between nerve cells allows the signal to diffuse quickly; this enables an immediate response, such as, Take your hand off the stove!


Synapse
Presynaptic cell
Neurotransmitter released into synapse
Neurotransmitter
Neurotransmitter attached to receptor
Enzyme that destroys neurotransmitter
Postsynaptic cell

The distance between the presynaptic cell and the postsynaptic cell—called the synaptic gap—is very small and allows for rapid diffusion of the neurotransmitter. Enzymes in the synapatic cleft degrade some types

of neurotransmitters to
terminate the signal.

When the neurotransmitter binds the receptor on the surface of the postsynaptic cell, the electrochemical potential of the target cell changes, and the next electrical impulse is launched. The neurotransmitters that are released into the chemical synapse are degraded quickly or get reabsorbed by the presynaptic cell so that the recipient nerve cell can recover quickly and be prepared to respond rapidly to the next synaptic signal.

Summary

Cells need to communicate to coordinate activities taking place both inside their own cytoplasm and with other cells. The former is called **intracellular communication** and the later **extracellular communication**. Intracellular communication often is triggered by an extracellular signal and is used to generate the response. The internal communication chain that connects the external signal to the internal response is called a **transduction pathway**, and the process can be summarized as the three stage process: **Signal -> Transduction -> Response**.

Communication between cells can be either **direct**, where their cytoplasm is connected by **gap junctions**, or **indirect**, where the signal is released into the extracellular fluid around the signaling cell. Indirect communication has four components: a **signaling cell**, a signal molecule (called a **ligand**), and a **target cell**, that has a **receptor protein** that responds to the ligand. Indirect communication can be differentiated by the proximity of the signal release to the target. **Paracrine communication** occurs when the signaling cell releases the signal molecule (called a **paracrine**) close to the target and the molecules reach the target receptors solely by diffusion through the interstitial fluid. **Hormonal communication** occurs when the signal molecule (called a **hormone**) diffuses into the blood stream and is distributed to interstitial fluid all around the body. Hormonal communication can target cells that are a great distance from the signalling cells and a given signal can affect many different types of cells simultaneously.

In addition to general paracrine communication, there are two special types of paracrine activity. **Autocrine communication** is used to describe signaling cells that have receptors for their own signaling molecules. They are both the signal initiator and the target cell. **Neuronal communication** refers to the special case of neurons. Like hormonal communication, neuronal communication can span long distances. Unlike hormonal communication: 1) the distance is spanned by physically extending a portion of the neuron itself (the axon) to the target cell, 2) the signal molecule (called a **neurotransmitter**) only diffuses through the interstitial fluid in the synapse, and 3) the target cell is very specific, even more specific than general paracrine targets.

Review Questions

Exercise:

Problem:

The secretion of hormones by the pituitary gland is an example of ________________.

a. autocrine signaling
b. paracrine signaling
c. endocrine signaling
d. direct signaling across gap junctions

Solution:

C

Exercise:

Problem:

Endocrine signals are transmitted more slowly than paracrine signals because ____________.

a. the ligands are transported through the bloodstream and travel greater distances
b. the target and signaling cells are close together
c. the ligands are degraded rapidly
d. the ligands don't bind to carrier proteins during transport

---

Solution:

A

Free Response

Exercise:

Problem:

What is the difference between intracellular signaling and intercellular signaling?

Solution:

Intracellular signaling occurs within a cell, and intercellular signaling occurs between cells.

Exercise:

Problem:

How are the effects of paracrine signaling limited to an area near the signaling cells?

Solution:

The secreted ligands are quickly removed by degradation or reabsorption into the cell so that they cannot travel far.

Types of Signal Molecules

This module provides an overview of the major classes of signaling molecules based on their polarity and chemical structure. It contains content from: Signaling Molecules and Cellular Receptors (m44451)

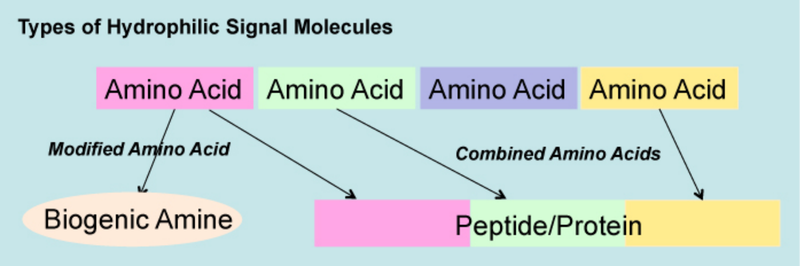
Ligands produced by signaling cells that subsequently bind to receptors in target cells act as chemical signals. These chemical signals travel to the target cells to coordinate responses. The types of ligands that serve as signal molecules are incredibly varied and range from small proteins to small ions like calcium (Ca^{2+}).

Signal molecules fall into a variety of different chemical classes. In general, all the signaling molecules in a given chemical class have similar characteristics in terms of polarity and receptor location (intracellular versus extracellular); however there are a few notable exceptions. For example, thyroid hormone belongs to the class of signal molecules known as bioamines. This class of molecule normally is hydrophilic attaching to cell-surface receptors; however thyroid hormone is an exception and is hydrophobic with intracellular receptors. An exception to the general receptor location rule has been found with several steroid molecules. Steroid receptors have long been identified as intracellular. However, recently a class of cell surface receptors for steroid molecules like estrogen has been identified and described.

Since the vast majority of signal molecules in a chemical class do have the same polarity, this section will begin by discussing the signal molecule chemical classes based on their polarity.

Hydrophilic Signaling Molecules

Water-soluble signal molecules are polar and therefore cannot pass through the non-polar center of the plasma membrane unaided. Sometimes, they are too large to pass through the membrane at all. Instead, most water-soluble ligands bind to the extracellular domain of cell-surface receptors. This group of signal molecules is quite diverse and most are based on amino acids. In some cases the unmodified amino acid is the signal molecule, in others the amino acid is modified into a derivative, and in still others amino acids are combined into peptides and proteins ([link]). Some additional small, polar, signal molecules also are hydrophilic.



The majority of hydrophilic signal molecules are based on amino acids. They may be unmodified amino acids, modified amino acids, or strings of amino acids (peptides or proteins).

Amino Acids

There are four amino acids that function as signal molecules. They are:

- Glutamate (made from glucose)
- Aspartate (made from glucose)
- Glycine (made from 3-phosphoglycerate)
- GABA (gamma-Aminobutyric Acid made from glutamate)

All four are neurotransmitters that are synthesized and stored in vesicles by the neurons that release them. Release is via exocytosis.

Biogenic Amines

All the members of this group are hydrophilic *except* thyroid hormone. The biogenic amine molecules all contain an amine group and are derived by modifying an amino acid. This group can be further differentiated by the specific amino acid that is modified to make the signal molecules.

- Catecholamines (derived from a single tyrosine molecule)
- Thyroid Hormone (derived from two tyrosine molecules)
- Histamine (derived from histadine)
- Serotonin (derived from tryptophan)

All are synthesized in the cytosol of the cells that release them and are stored in vesicles. They are released through exocytosis. The biogenic amines a cell is able to produce is determined by the specific enzymes that are active in the synthesizing cell.

Catecholamines

Catecholamines are derived from a single tyrosine molecule. The best known examples of this group are:

- Dopamine
- Norepinephrine (Noradrenalin)
- Epinephrine (Adrenalin)

All three are neurotransmitters. Norepinephrine is the primary neurotransmitter of the postsynaptic neurons in the sympathetic nervous system. Epinephrine is best known as the sympathetic system hormone adrenalin, and while dopamine also is released with epinephrine from the adrenal medulla, it is better known for its functions as an inhibitor of antagonist muscle action when released from substantia nigra neurons and as the reward or pleasure neurotransmitter when released from VTA neurons.

Thyroid Hormone

Thyroid Hormone is derived from two tyrosine molecules. Thyroid hormone molecules differ from others in the class by being hydrophobic rather than hydrophilic, and like most other hydrophobic signals its mode of action is intracellular. Even though thyroid hormone is hydrophobic and has intracellular receptors the molecules do not passively diffuse across the plasma membrane but must be actively transported across to reach their receptors. Thyroid hormone can act on almost all cells in the body and is involved in setting the metabolic rate and in cell growth and division.

Histamine

Histamine is derived from the amino acid histadine. It is best know for its paracrine signaling function in the immune response which causes inflammation and the symptoms of many allergies. However, it also is an important signal molecule involved in both GI signaling and as a neurotransmitter. In the GI tract histamine acts as an enteroendocrine hormone that stimulates the release of gastric acid. As a neurotransmitter histamine release from the hypothalamus increases wakefulness and prevents sleep. It also has been implicated in a number of other neural functions including schizophrenia and proper libido and erectile function in males.

Seratonin

Serotonin is derived from the amino acid tryptophan. It is found in the GI tract, blood platelets, and the central nervous system. Serotonin is involved in a variety of important functions. In the GI tract and blood platelets the general function is to cause smooth muscle contraction. However, the most publicized function of serotonin is its association with general mood and mood disorders in the central nervous system.

The connection of serotonin to feelings of well being, or lack of well being (like depression and anxiety), has led to the marketing of a number of psycopharmaceutical drugs designed to increase serotonin levels in the brain. The most well known are the Selective Serotonin Reuptake Inhibitors

(SSRIs) that block the reuptake channels in the presynaptic cells thereby increasing the time and concentration of serotonin in the synapse. This pathway is the one targeted by the most common antidepressant drugs like Prozac and Zoloft.

Peptides and Proteins

The majority of signal molecules are peptides and small proteins. Both are made of a chain of amino acids, the difference being the number of amino acids that make up the molecule. Generally, if the signal molecules has fewer than 50 amino acids it is considered a peptide. If it has more it is considered a protein. All interact with extracellular cell-membrane receptor proteins.

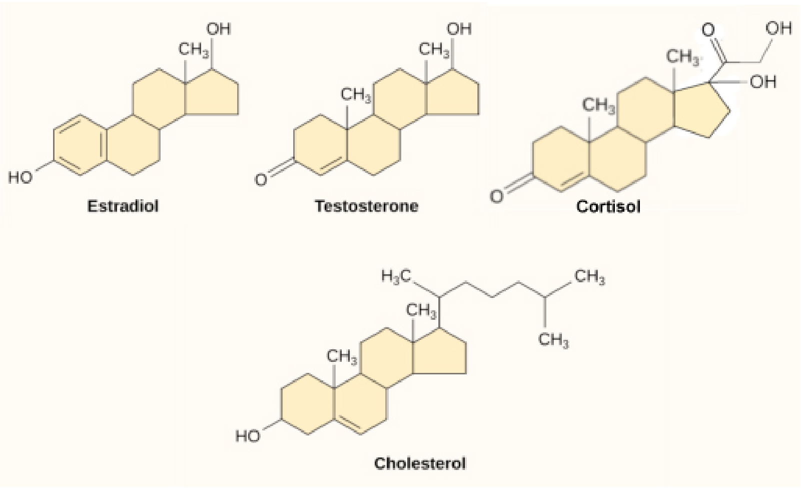
The peptides and proteins are synthesized as inactive larger molecules called *prepropeptides* in the rough ER of the cell that releases them. A portion of this molecule is cleaved off by an enzyme leaving a still inactive *propeptide* that is released in a transport vesicle and may undergo additional processing in the Golgi apparatus. The final version of the propeptide is stored in a vesicle and activated when another enzyme cleaves additional amino acids from the molecule before it is released by exocytosis. Like the previously discussed hydrophilic signal molecules, peptide and protein signal molecules are synthesized by the cell that releases them and the specific signal molecule synthesized depends on the enzymes and transcription factors that are active in the synthesizing cell.

Small Hydrophobic Signaling Molecules

Small hydrophobic ligands can directly diffuse through the plasma membrane and interact with internal receptors. Two important members of this class of ligands are the steroid hormones and eicosanoids. Other hydrophobic hormones include thyroid hormones and vitamin D. In order to be soluble in blood, hydrophobic ligands must bind to hydrophilic carrier proteins while they are being transported through the bloodstream.

Steroids

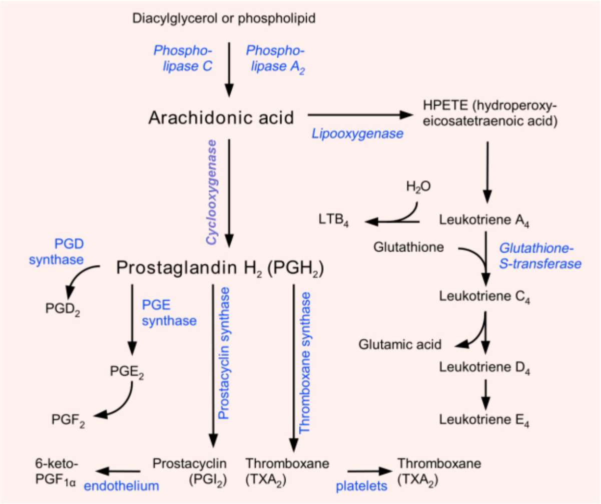
Steroids are lipids that have a hydrocarbon skeleton with four fused rings; different steroids have different functional groups attached to the carbon skeleton. Cholesterol is the precursor of steroid hormones and an important structural component of biological membranes ([link]). Steroid hormones include the female sex hormone, estradiol, which is a type of estrogen; the male sex hormone, testosterone; and the stress hormone, cortisol, released by cells in the adrenal cortex. They are not stored, but are synthesized on demand from cholesterol in the plasma membrane.



Steroid hormones all retain the central four-carbon ring of their precursor, cholesterol but have different modifications to the functional groups. Because these molecules are small and hydrophobic, they can diffuse directly across the plasma membrane into the cell, where they interact with internal receptors, although some also bind to cell-surface receptors.

Eicosanoids

Eicosandoids also are synthesized on demand from a phospholipid called, *arachidonic acid*, that is found in the plasma membrane. There are two major pathways that use the arachidonic acid as a precursor: the **cyclooxygenase pathway** and the **lipoxygenase pathway**.



Eicosanoids begin as the plasma membrane phospholipids and that are modified by two different pathways. The cyclooxygenase pathway creates prostaglanins, prostacyclins, and thromboxanes. The lipoxygenase pathway creates leukotrienes. (Modified from Jfdwolff figure on Wikimedia Commons.)

Cyclooxygenase pathway: results in three classes of molecules depending on the specific enzymes activated. these are prostaglandins, prostacyclins, and thromboxanes **Prostaglandins** can have hormone-like functions, but they are produced by many cells around the body rather than specific organs like most other hormones. Their effects also can be dramatically different in different tissue types. **Prostacyclins** act as paracrines and are synthesized by cells in blood vessel walls and are potent vasodilators. They also inhibit platelet aggregation and act as an anti-clotting agent in healthy endothelium. **Thromboxanes**, also paracrines, are rleased by platelets and have the reverse effect to prostacyclins. They are potent vasoconstrictors and promote platelet aggregation. The cyclooxygenase pathway is often abbreviated as COX, and an important class of analgesic and anti-inflammatory drugs called *COX-inhibitors* target this pathway. The COX pathway inhibitors also go by the name *Non-Steroidal Anti-Inflammatory drugs* or *NSAIDs* and include aspirin, ibuprofen (Advil, Motrin), naproxen (Aleve), and a host of other drugs that target different aspects of the pathway.

Lipoxygenase pathway: produces a class of signal molecules called leukotrienes, because they are produced by certain types of leukocytes (white blood cells). They function as paracrines and trigger inflammation. They often are released with both histamines and prostaglandins, which also trigger inflammation. In addition, leukotrienes trigger contractions in the smooth muscle lining the bronchioles and therefore play a significant role in creating the symptoms of asthma. There are asthma drugs on the market that target leukotrienes by either inhibiting the the lipoxygenase pathway or by blocking the leukotriene receptors.

Other Signaling Molecules

Calcium ions (Ca^{2+}) control many diverse and critical cellular processes ranging from muscle contraction to exocytosis to gene transcription to apoptosis. Many of the signaling activities are intracellular acting as second messengers, although they also are one of the ions used for direct intercellular communication through gap junctions. Numerous proteins in intracellular signal transduction pathways are modulated directly or indirectly by calcium. Prominent examples include kinases and phosphatases, transcription factors, and calmodulin. Because of their critical role, calcium levels must be tightly controlled both inside and outside the cell.

Nitric oxide (NO) is a gas that also acts as a signaling molecule. It is able to diffuse directly across the plasma membrane, and one of its roles is to interact with receptors in smooth muscle and induce relaxation of the tissue. NO has a very short half-life and therefore only functions over short distances. Nitroglycerin, a treatment for heart disease, acts by triggering the release of NO, which causes blood vessels to dilate (expand), thus restoring blood flow to the heart. NO has become better known recently because the pathway that it affects is targeted by prescription medications for erectile dysfunction, such as Viagra (erection involves dilated blood vessels).

Non-coding RNA (ncRNA)molecules have been shown to have both intercellular signaling and catalytic functions. They have additional capabilities not found in other types of signaling molecules in that they do not just rely on the transduction pathways of the receiving cell, but can carry encoded information in the signal. As such they can modify the splicing and translation of mRNA and they have been shown to play roles in the regulation and structure of many cellular processes.

Exogenous Signal Molecules

Signal molecules are not just generated by our cells. They naturally occur all around us and exogenous signals regularly enter our bodies through our gastrointestinal and respiratory tracts. Because cells of all species use many of the same signal molecules these molecules often induce a response in our body that can affect both our physical function and our behavior. They can be beneficial or deadly, but they all work by activating or blocking existing receptors in our cells.

When the signal molecule is produced by our body we call it an **endogenous** signal. When the signal molecule is produced by an external organism, or artificially, we call it an **exogenous** signal. Many receptors are named for the exogenous signals that were used to identify them, for example nicotinic (from tobacco), muscarinic (from mushrooms), cannabinoid (from cannabis). They all have endogenous counterparts.

Exogenous signal molecules can be divided into the four major sources:

- Bacterial
- Multicellular tissue
- Pheromones
- Synthetic molecules

Bacteria (as well as parasites) produce and release a variety of signal molecules in the GI tract that can be absorbed and cause both physical and behavior changes in the host. When the changes are beneficial we call the organism a **probiotic**. When they are detrimental they are defined as a disease or disorder. Recognizing bacteria as an important source of beneficial as well as detrimental signal molecules is fairly recent and is undergoing intense focus from a number of angles. New effects both physical and behavioral continue to be identified and explored through research studies. While many may be beneficial on the borders of the body, if the bacteria (or parasite) manages to get into the interior of our body (blood and other body fluids) it will cause disease.

Multicellular tissue signals, depending on the type, can enter the body via several different pathways. They can be ingested and absorbed through the intestines; they can be inhaled and absorbed in the lungs, or they can be injected directly into the tissue or bloodstream. Many are actively pursued by the receiving organism because of the signal's effect. For example caffeine is purposefully consumed by people all over the world. The same is true of nicotine, THC and CBD (in cannabis), psilocybin (in mushrooms), mescalin (in cactus), opium (in poppies), and a host of other naturally produced substances. Other signal molecules from multicellular tissue enter the body without conscious participation, for example the anticoagulant paracrines in the saliva of mosquitoes and other blood feeding insects that cause the swelling and itching from their bites.

Pheromones are signal molecules secreted or excreted into the environment with the express function of affecting the behavior of other individuals of the same species. Pheromones often are involved in mediating territorial and sexual behaviors. This form of social communication has been well studied in a variety of animals from insects to mammals. For example, the urine of dogs contains pheromones that are used to mark territory. By smelling the pheromones, other dogs can not only tell which dog or dogs have marked the spot, but how long ago each dog was there, and even aspects of the physiological states of the dogs that marked it. There is evidence that humans, like other mammals, also produce pheromones, especially from apocrine glands around the nipples, in the groin area, and in the axillary regions. However, the complexity of chemicals from these glands has made it difficult to isolate any specific molecule as a definite human pheromone signal molecule.

Synthetic molecules are molecules that are not derived or copied directly from naturally produced signal molecules, but have bioactive signaling capabilities. Many of our synthetic materials, particularly plastics and pesticides unintentionally have released molecules that mimic naturally occurring signal molecules. Many can be shown to enter the body although not all the effects are easily seen. Some are known to mimic estrogen and have been blamed for both physical and behavioral changes in humans. Others have been explicitly designed for psychotropic effects both medicinal and illegal.

Section Summary

Signal molecules are ligands that bind to receptor proteins and stimulate an action by the cell. They can be divided into two categories based on the polarity of the molecule, **hydrophilic** and **hydrophobic**. *In general*, hydrophilic signal molecules interact with receptors on the surface of the plasma membrane and hydrophobic signal molecules interact with receptors inside the cell that regulate cellular transcription.

Hydrophilic signal molecules are generated by the cells that release them and are stored in vesicles until release. Most are either **amino acids**, **biogenic amines** (derivatives of amino acids), or **peptides/proteins** (strings of amino acids). Biogenic amines are classified by the amino acid from which they are derived: *catecholamines* (single tyrosine), *thyroid hormone* (two tyrosines), *histamine* (histadine), and *serotonin* (tryptophan). Some additional small, polar molecules not based on amino acids also act as hydrophilic signal molecules.

Hydrophobic signal molecules also are generated by the cells that release them, but are generated on demand from plasma membrane lipids rather than stored in vesicles. **Steroids** are derived from *cholesterol* and include the sex hormones and the stress hormone cortisol. **Eicosanoids** are derived from the membrane phospholipid, *arachidonic acid*. They fall into two classes depending on the initial enzyme used to modify the arachidonic acid. The **cyclooxygenase (COX) pathway** gives rise to the *prostaglandins*, *prostacyclins*, and *thromboxanes*. The **lipoxygenase pathway** gives rise to the *leukotrienes*.

There are many **additional endogenous molecules** that act as signaling molecules. These include *calcium ions* (Ca^{2+}), *nitric oxide (NO)*, and *non-coding RNA molecules (ncRNA)*. **Exogenous molecules** also can interact with cellular receptors and affect physical and mental activities. These include molecules release by *bacteria*, *multicellular tissue*, *pheromones*, and *synthetic chemicals*.

Signal Molecule Release, Transport, and Binding
This section covers the fate of signal molecules from release to receptor binding.

Signal Molecule Transport and Persistence

All indirect signal molecules diffuse through interstitial fluid. Paracrines, autocrines, neurotransmitters, and most cytokines have close targets that they reach entirely by diffusion. The responses they trigger are localized both because the signal molecule concentration decreases with distance from the source and because these signal molecules are quickly degraded or removed from the interstitial fluid.

Hormones, neurohormones, and some cytokines diffuse through the interstitial fluid and into blood capillaries near the source cells in high enough concentrations to affect distant cells. They then travel in the blood plasma long distances from the releasing cells to their target receptors. These signal molecules must be able to travel with the polar water molecules that make up most of the blood plasma. The hydrophilic (polar) signal molecules dissolve directly in the plasma. The hydrophobic (nonpolar) signal molecules must bind to polar, carrier-proteins to travel in the blood plasma.

The time a specific class of signal molecules remains active in the blood is measured by its **half-life**. This is the time it takes a given concentration to decrease by half. For example, if the initial concentration is 10 mg (milligrams) per deciliter (tenth of a liter) of blood plasma, then the half-life is the time it takes for the concentration to decrease to 5 mg per deciliter. If this time was 10 minutes, then in the next ten minutes the concentration would decrease by half again to 2.5 mg per deciliter. Over the next 10 minutes it would decrease by half again to 1.25 mg per deciliter, and so on.

Different signal molecules have different half-lives depending on how they are removed from circulation. Some are broken down by enzymes; others are filtered out of the blood by the kidneys. Signal molecules that dissolve directly in the blood tend to have shorter half-lives than signal molecules that are bound to a carrier protein. For example, insulin (a polar, peptide hormone) has a half-life of less than 10 minutes, while cortisol (a nonpolar, steroid hormone) has a half-life that exceeds 90 minutes.

Binding Speed and Response

A signal molecule can only bind to receptors for which it has an affinity, receptors that match its shape and charge pattern. The concept of affinity was discussed earlier in the *Energy and Metabolism* chapter in the section on enzymes. The same chemical processes found in enzyme binding also operate in signal-molecule binding.

Receptor/signal binding is reversible and generally only lasts for a short time. Signal molecules often will bind to more than one receptor type; however, not all receptor types have the same affinity for a given signal molecule. For this reason more of a given signal molecule will bind to the receptors for which it has the highest affinity first. As the higher affinity receptors become saturated, the signal molecule will bind more frequently to the available lower affinity receptors.

Cells do not have just one receptor, nor do they have a single receptor type. Their surfaces are covered with millions of proteins that carry out different functions. This means that the target cells for a specific signal molecule also may have many other types of receptors that bind to different signal molecules. The available receptor types differ greatly between cell types and determine the differential actions of the cell type.

The strength and speed of a given cell's response to a specific signal-molecule depends on three factors:

- The affinity of the cell's receptors for the signal molecule *(The higher the affinity, the faster and stronger the response.)*
- The concentration of the signal molecule *(The higher the concentration, the faster and stronger the response.)*
- The number of receptors for that signal molecule the cell has on its surface *(The more receptors, the faster and stronger the response.)*

The number of receptors of a given type available at a specific cell varies and is dynamically controlled.

Receptor Regulation

Down Regulation describes a response in a target cell that causes it to decrease its sensitivity to an external variable by reducing the number of receptors available at that cell. This generally is an internal negative feedback response to high concentrations of signal molecule continuously stimulating the cell. You might think of this as the cellular equivalent of people plugging their ears to help them concentrate in a noisy room. **Up Regulation** is the opposite of down regulation, where the cell increases the number of receptors in response to low concentrations of signal molecule providing less than adequate stimulation.

Agonist and Antagonist are terms you likely encountered in anatomy when you studied muscles. An agonist is a muscle that acts to change the position of a bone, while an antagonist is a muscle that actively opposes that action by pulling in the opposite direction. The same concept is found in the relationships of signal molecules and receptors. An **agonist** is a signal molecule that attaches to a receptor and activates the receptor molecule response. An **antagonist** is a signal molecule that attaches to the receptor and does not activate the receptor molecule response. Antagonists, therefore, actively block or dampen a response by competing with the agonists for available receptors.

Note:

Connection to Addiction

Exogenous signal molecules act as either agonists or antagonists when binding to target cell receptors. High concentrations of an agonist or antagonist are sometimes introduced and maintained in the body over a period of time either for prescribed therapeutic effects or individually motivated euphoric effects. The resulting over or under stimulation can cause the target cells to up-regulate or down-regulate in response. This is why the dosage of some agents must be increased over time to maintain the desired effect.

If the agonist or antagonist use is continued for a period of time and then suddenly stopped it can cause disequilibrium in the target cells until normal sensitivity is re-established by compensating up-regulation or down-regulation. These transitional changes in cellular sensitivity often cause withdrawal symptoms that can have profound physiological effects up to and including death.

Additionally a cell that has been over stimulated for a period of time can over compensate when up-regulating during withdrawal and make the cell more sensitive than its original state. If an addict relapses (resumes using the agonist), their initial response after relapse may be even more intense than the first time they encountered the agonist. This is why breaking an addiction is so difficult.

Section Summary

All indirect signal molecules are released into the interstitial fluid around the signaling cell. Some affect only nearby cells within diffusion range. Others diffuse into the blood stream and affect target cells throughout the body. The time a signal molecule persists in the blood stream is measured as **half-life**. Half-life is the time it takes for the signal-molecules to reach a concentration that is half of the starting concentration.

A given type of signal-molecule may bind to more than one receptor type, but generally has a higher affinity for some receptor types over others. Likewise, a given receptor may bind to more than one type of signal molecule, but may have different affinities for one signal-molecule type over others. Target cells have millions of proteins and so may contain many different types of receptors. The strength and speed of binding is affected by three major factors: the affinity of the signal-molecule and the receptor, the concentration of the signal molecule, and the number of receptors on the target cell.

The response of a target cell is partially controlled by the conditions at the cell, including dynamic control of the number of receptors available by processes in the target cell. An over-stimulated cell may **down-regulate** the number of receptors by removing or inactivating them. An under-stimulated cell may **up-regulate** the number of receptors by adding or activating more. Responses also may be increased or decreased at the target cell by the presence of additional agonist or antagonist molecules. The effects of artificially introduced agonists or antagonists on down-regulation and up-regulation processes in the cell explain the withdrawal effects in addiction.

Intracellular Signal Transduction

By the end of this section, you will be able to:

- Compare internal receptors with cell-surface receptors
- Recognize the general relationship between a ligand’s structure and its mechanism of action
- Describe the four possible combinations of activation mechanisms and modes of action for transduction via cell-surface receptors
- Describe the effects of second messengers in signal transduction

In an earlier section we discussed types of signal molecules. While the characteristics of the signal molecule are important for understanding how the molecule gets to the target cells and the ways it is able to interact with those cells, the actual response (interpretation) depends on the specific transduction pathways activated in the responding cell.

Transduction pathways are the way in which target cells convert a signal (or message) into action. That action (response) is the target cell's *interpretation* of the message. Just like different individuals may have vastly different interpretations of the same message, so different cell types often have different responses to the same signal molecule. The characteristics of the target cell determine the specific response to a signal molecule. The signal molecule can only turn the response that already exists in a specific target cell on or off.

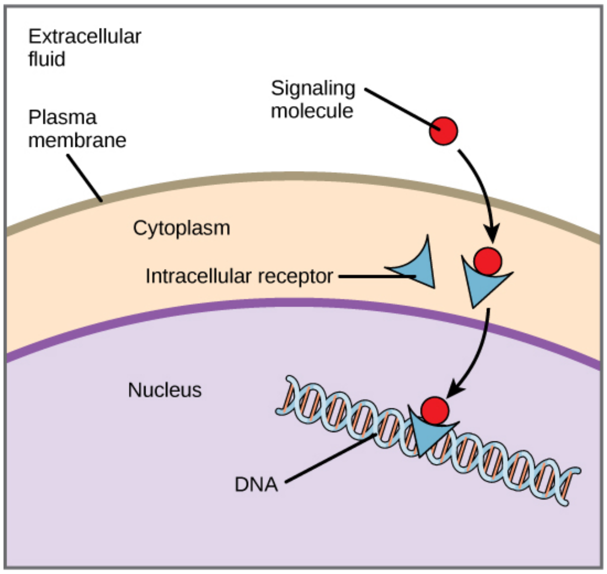
In the chapter on *Energy and Metabolism* we discussed characteristics of enzymes and substrates. The same binding principles are at play with all ligand/receptor interactions, including signal-molecule/receptor binding. So if you do not remember them, this would be a good time to review the concepts of *induced fit* and *affinity*. The receptor proteins that trigger different transduction pathways for the same signal molecule also often have slight differences in the receptor binding sites that affect the induced fit and affinity for the signal molecule. This can affect the whole body response to specific signals, but is particularly important in how certain drugs that act as agonists or antagonists for one of these responses can cause unwanted side effects (different responses) from other cells with similar receptors for the same signaling molecule.

Activation Mechanisms and Modes of Action

Ligands are either hydrophilic (polar) or hydrophobic (non-polar). Hydrophilic ligands cannot cross the non-polar interior of the plasma membrane, so they must bind to cell-surface receptors on the outside of the plasma membrane and effect their changes via the receptor protein. Because hydrophobic signal molecules can easily diffuse through the plasma membrane, they can attach to receptor proteins in the cytoplasm inside the cell. However, it is important to remember that just because a signal molecule *can* activate internal receptors does not mean that all its receptors *are* internal. Some hydrophobic signal molecules bind to cell-surface receptors as well as intracellular receptors in certain cells.

Intracellular Receptors and Signal Transduction

Internal receptors, also known as intracellular or cytoplasmic receptors, are found in the cytoplasm of the cell and respond to hydrophobic ligand molecules that are able to travel across the plasma membrane. Once inside the cell, many of these molecules bind to receptor sites on proteins that act as regulators of mRNA synthesis (transcription) to mediate gene expression. These are the transcription factors discussed in the earlier chapter: *DNA and Protein Synthesis*. When the ligand binds to the internal receptor, a conformational change is triggered that exposes a DNA-binding site on the protein. The ligand-receptor complex moves into the nucleus, then binds to specific regulatory regions of the chromosomal DNA and promotes the initiation of transcription ([link]). Internal receptors directly influence gene expression without having to pass the signal on to other receptors or messengers.



Hydrophobic signaling molecules typically diffuse across the plasma membrane and interact with intracellular receptors in the cytoplasm. Many intracellular receptors are transcription factors that interact with DNA in the nucleus and regulate gene expression.

Cell-Surface Receptors and Signal Transduction

Cell-surface receptors, also known as transmembrane receptors, are cell surface, membrane-anchored (integral) proteins that bind to hydrophilic ligand molecules on the external surface of the plasma membrane. This type of receptor spans the plasma membrane and performs signal transduction, in which an extracellular signal is converted into an intracellular signal and response. Ligands that interact with cell-surface receptors do not have to enter the cell that they affect. Cell-surface receptors are also called cell-specific proteins or markers because they are specific to individual cell types.

Because cell-surface receptor proteins are fundamental to normal cell functioning, it should come as no surprise that a malfunction in any one of these proteins could have severe consequences. Errors in the protein structures of certain receptor molecules have been shown to play a role in hypertension (high blood pressure), asthma, heart disease, and cancer.

Each cell-surface receptor protein has three main components: an external ligand-binding domain, a hydrophobic membrane-spanning region, and an intracellular domain inside the cell. The ligand-binding domain is also called the **extracellular domain**. The size and extent of each of these domains vary widely depending on the specific type of receptor protein.

Note:

Evolution Connection

How Viruses Recognize a Host

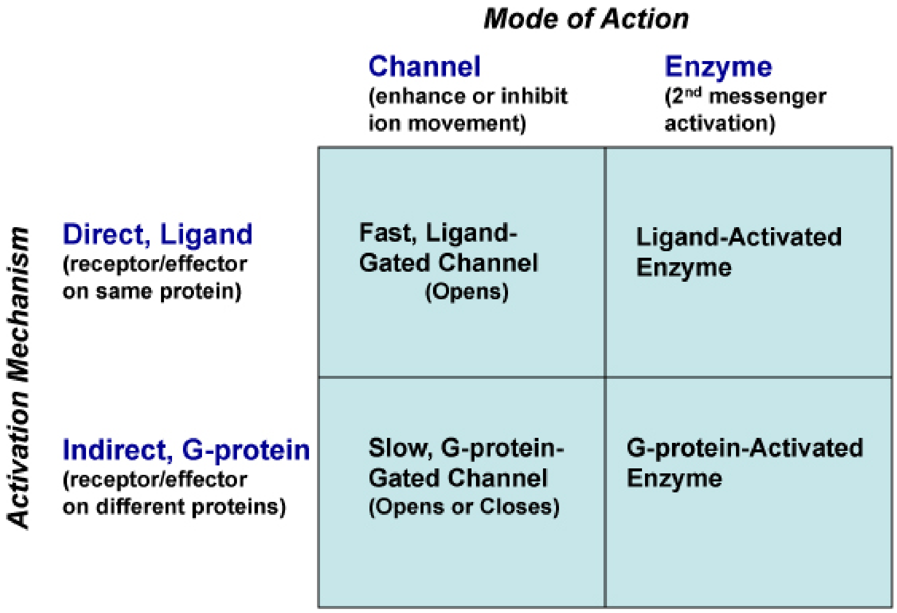
Unlike living cells, many viruses do not have a plasma membrane or any of the structures necessary to copy themselves. Some viruses are simply composed of an inert protein shell surrounding DNA or RNA. To reproduce, viruses must invade a living cell, which serves as a host, and then take over the hosts cellular apparatus. But how does a virus recognize its host?

Viruses often bind to cell-surface receptors on the host cell. For example, the virus that causes human influenza (flu) binds specifically to receptors on membranes of cells of the respiratory system. Chemical differences in the cell-surface receptors among hosts mean that a virus that infects a specific species (for example, humans) cannot infect another species (for example, chickens).

However, viruses have very small amounts of DNA or RNA compared to humans, and, as a result, viral reproduction can occur rapidly. Viral reproduction invariably produces errors that can lead to changes in newly produced viruses; these changes mean that the viral proteins that interact with cell-surface receptors may evolve in such a way that they can bind to receptors in a new host. Such changes happen randomly and quite often in the reproductive cycle of a virus, but the changes only matter if a virus with new binding properties comes into contact with a suitable host. In the case of influenza, this situation can occur in settings where animals and people are in close contact, such as poultry and swine farms.[footnote] Once a virus jumps to a new host, it can spread quickly. Scientists watch newly appearing viruses (called emerging viruses) closely in the hope that such monitoring can reduce the likelihood of global viral epidemics.

A. B. Sigalov, The School of Nature. IV. Learning from Viruses, *Self/Nonself* 1, no. 4 (2010): 282-298. Y. Cao, X. Koh, L. Dong, X. Du, A. Wu, X. Ding, H. Deng, Y. Shu, J. Chen, T. Jiang, Rapid Estimation of Binding Activity of Influenza Virus Hemagglutinin to Human and Avian Receptors, *PLoS One* 6, no. 4 (2011): e18664.

Cell-surface receptors are involved in most of the signaling in multicellular organisms. There are two general categories of cell-surface signal transduction based on the mechanism used to activate the transduction process. In **direct activation** (or **ligand activation**) the receptor and effector share the same protein, so the ligand directly activates the transduction process. With **indirect activation** the receptor is on one protein and the effector is on a different, physically separated, protein. This requires the proteins to communicate to activate the effector. The intermediate signal is a G-protein so this mechanism is also called **G-protein-linked activation**. There also are two modes of effector action: **channels** and **enzymes**. Therefore there are four possible combinations of activation types and modes of action, see ([link]).

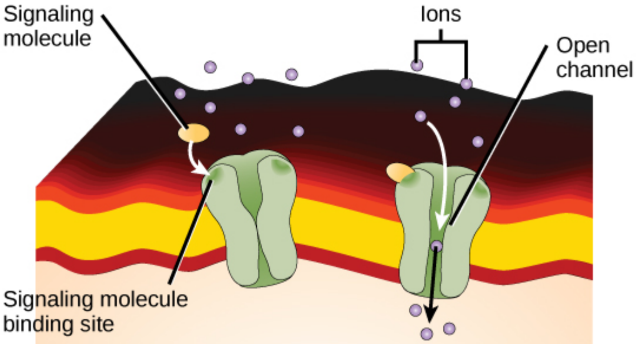


Mode of Action

Activation Mechanism	**Channel** (enhance or inhibit ion movement)	**Enzyme** (2^{nd} messenger activation)
Direct, Ligand (receptor/effector on same protein)	Fast, Ligand-Gated Channel (Opens)	Ligand-Activated Enzyme
Indirect, G-protein (receptor/effector on different proteins)	Slow, G-protein-Gated Channel (Opens or Closes)	G-protein-Activated Enzyme

The four possible combinations of activation mechanisms and modes of action.

Direct, Ligand-Gated Channels are proteins with both a channel and a receptor. When a ligand binds to the receptor the channel opens through the membrane allowing specific ions to pass through. To form a channel the protein has an extensive membrane-spanning region. In order to interact with the phospholipid fatty acid tails that form the center of the plasma membrane, many of the amino acids in the membrane-spanning region are hydrophobic in nature. Conversely, the amino acids that line the inside of the channel are hydrophilic to allow for the passage of water or ions. When a ligand binds to the extracellular region of the channel, there is a conformational change in the protein's structure that allows ions such as sodium, calcium, magnesium, and hydrogen to pass through ([link]).

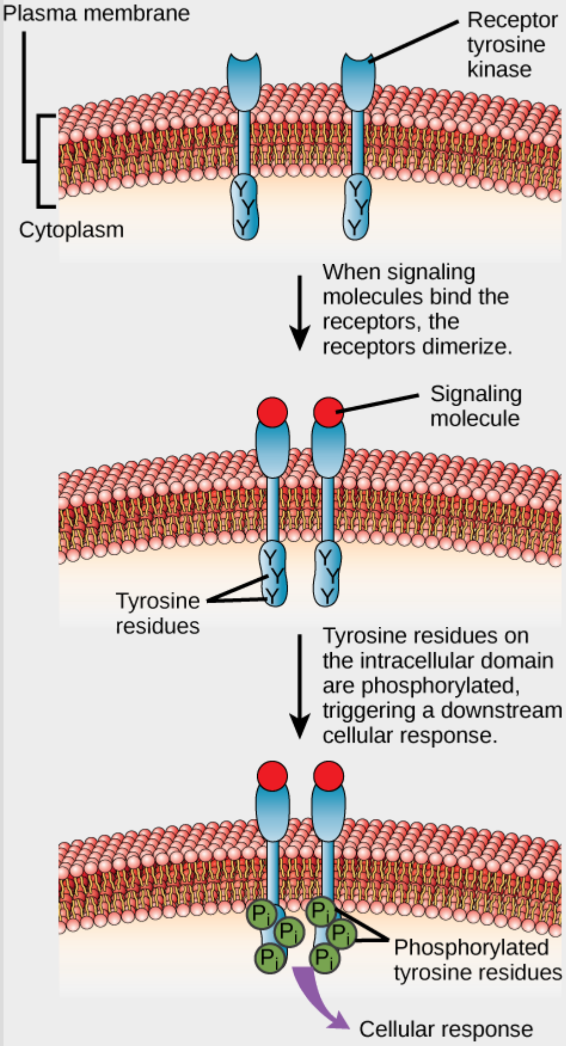


Ligand-gated ion channels form a pore through the plasma membrane that opens when the signaling molecule binds. The open pore then allows ions to flow into or out of the cell.

Because of their direct action, ligand-gated channels are fast to open and also fast to close. Their mode of action is to open the channel rather than close it.

Direct, Ligand-activated enzymes are cell-surface receptors with intracellular domains that are associated with an enzyme. In some ligand-activated enzymes, the intracellular domain of the receptor protein itself is an enzyme. Other ligand-activated enzymes have a small intracellular domain that directly interacts with the enzyme. The ligand-activated enzymes normally have large extracellular and intracellular domains, but the membrane-spanning region consists of a single alpha-helical region of the peptide strand. When a ligand binds to the extracellular domain, a signal is transferred through the membrane, activating the enzyme. Activation of the enzyme sets off a chain of events within the cell that eventually leads to a response. One example of this type of ligand-activated enzyme is the tyrosine kinase receptor protein ([link]). A kinase is an enzyme that transfers phosphate groups from ATP to another protein. When ligands bind to the tyrosine kinase receptor it results in the transfer of phosphate groups to the intracellular tyrosine molecules (tyrosine residues). First, signaling molecules bind to the extracellular domain of two nearby tyrosine kinase receptors. The two neighboring receptors then bond together, or dimerize. Phosphates are then added to tyrosine residues on the intracellular domain of the receptors (phosphorylation). The phosphorylated residues can then transmit the signal to the next messenger within the cytoplasm.

Note:


Plasma membrane
Receptor tyrosine kinase
Cytoplasm
When signaling molecules bind the receptors, the receptors dimerize.
Signaling molecule
Tyrosine residues
Tyrosine residues on the intracellular domain are phosphorylated, triggering a downstream cellular response.
Phosphorylated tyrosine residues
Cellular response

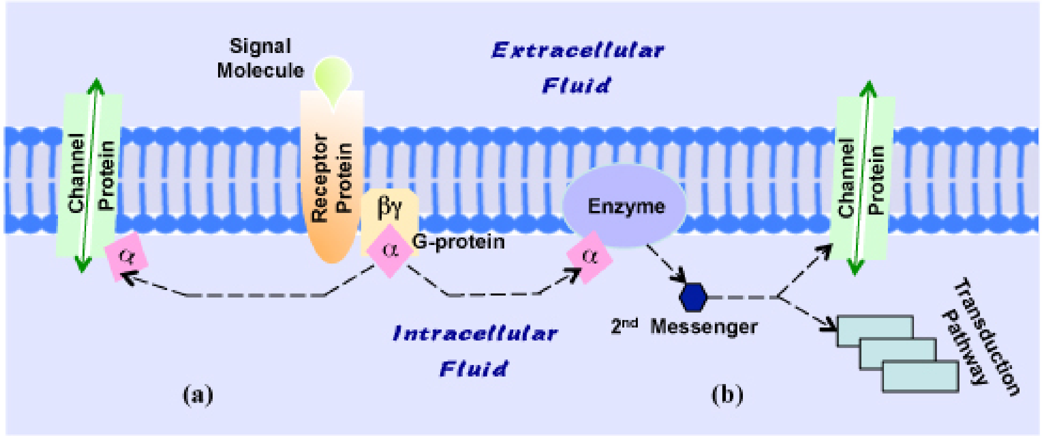
A receptor tyrosine kinase is an enzyme-linked receptor with a single transmembrane region, and extracellular and intracellular domains. Binding of a signaling molecule to the extracellular domain causes the receptor to dimerize. Tyrosine residues on the intracellular domain are then autophosphorylated, triggering a downstream cellular response. The signal is terminated by a phosphatase that removes the phosphates from the phosphotyrosine residues.

HER2 is a tyrosine kinase receptor. In 30 percent of human breast cancers, HER2 is permanently activated, resulting in unregulated cell division. Lapatinib, a drug used to treat breast cancer, inhibits HER2 receptor tyrosine kinase autophosphorylation (the process by which the receptor adds phosphates onto itself), thus reducing tumor growth by 50 percent. Besides autophosphorylation, which of the following steps would be inhibited by Lapatinib?

a. Signaling molecule binding, dimerization, and the downstream cellular response
b. Dimerization, and the downstream cellular response
c. The downstream cellular response
d. Phosphatase activity, dimerization, and the downsteam cellular response

Indirect, G-protein activation: a ligand binds to a surface receptor on a protein which activates an associated membrane protein called a G-protein. The activated G-protein then interacts with specific ion channels or enzymes in the membrane ([link]). When the G-protein is activated by a ligand the G-protein portion disassociates into two subunits (*α* and *βγ*). One or both of the G-protein subunits that dissociate are available to modify **G-**

protein-gated channels or **G-protein-activated enzymes** turning them on or off. When the *α* subunit is hydrolyzed back into GDP and the *βγ* subunit is deactivated they dissociate from the effector proteins they modified returning them to their original state and the subunits reassociate as the inactive G-protein.

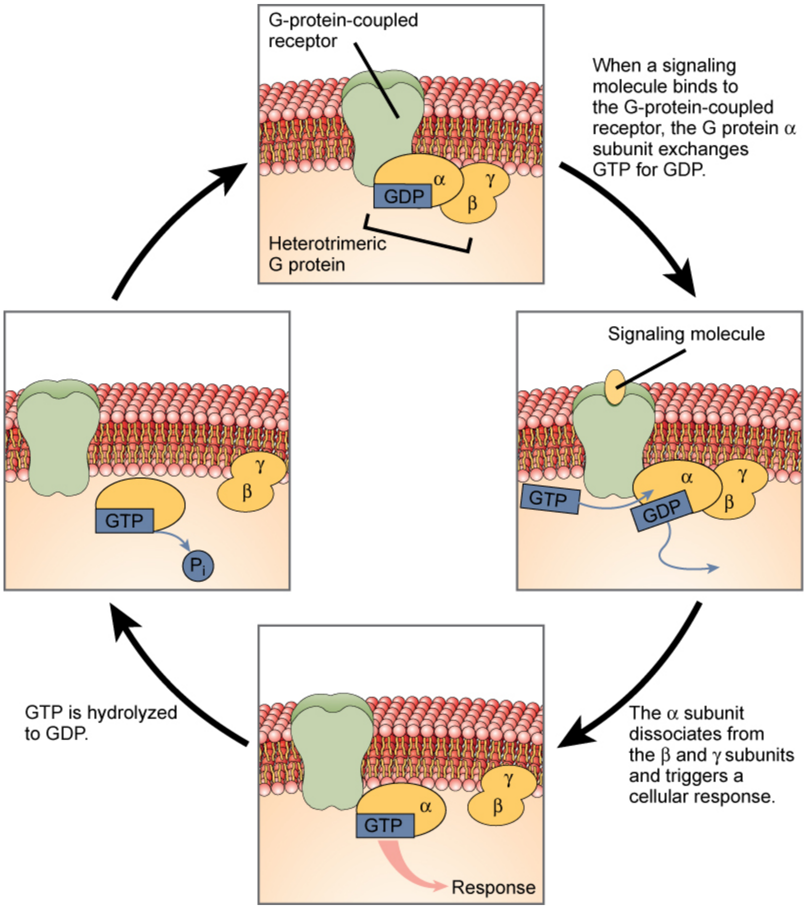


G-protein Activation: Either or both the alpha and beta subunits can activate intracellular proteins. The center shows the receptor and G-proteins. The left side (a) shows the alpha subunit opening a transmembrane channel. The right side (b) shows the alpha subunit activating an enzyme that then creates a second messenger.

Both G-protein-activated and direct, ligand-activated enzymes often catalyze the creation of small molecules like cAMP and cGMP called **Second Messengers**. An activated enzyme can create many copies of a second messenger. Each second messenger can activate many copies of subsequent proteins. In this manner a second messenger can greatly **amplify the signal** of the first messenger that attached to the extracellular receptor. The multiple copies of the second messenger also can trigger activation of many different intracellular proteins and pathways and **diversify the effect** of the first messenger. In this way, second messengers allow a single first messenger attaching to the external receptor to have both a large and diversified intracellular effect.

All G-protein-linked receptors have seven transmembrane domains, so the general mode of communication between the receptor and effector proteins is the same for a channel or an enzyme. However, each receptor has its own specific extracellular domain and G-protein-binding site and the effector proteins also vary in their binding sites and specific actions.

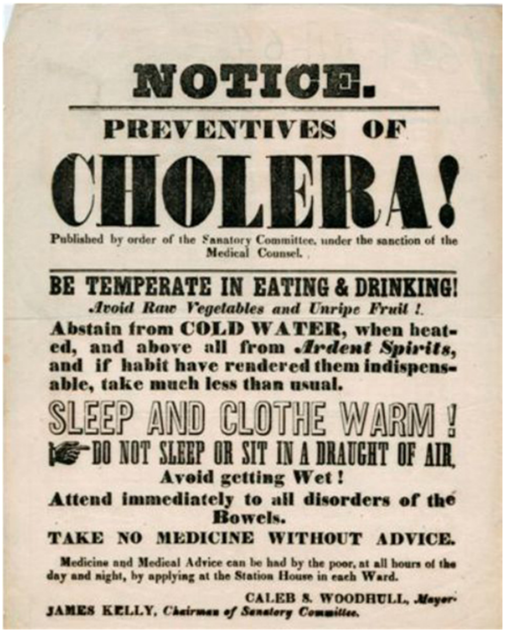
Cell signaling using G-protein-linked receptors occurs as a cyclic series of events ([link]). Before the ligand binds, the inactive G-protein can bind to a newly revealed site on the receptor specific for its binding. Once the ligand binds to the receptor, the resultant shape change activates the G-protein, which releases GDP and picks up GTP. The subunits of the G-protein then split into the *α* subunit and the *βγ* subunit. The activity terminates when the GTP on the active *α* subunit of the G-protein is hydrolyzed to GDP and the *βγ* subunit is deactivated. The subunits reassociate to form the inactive G-protein and the cycle begins anew.


G-protein-coupled receptor
GDP
α
γ
β
Heterotrimeric G protein
When a signaling molecule binds to the G-protein-coupled receptor, the G protein α subunit exchanges GTP for GDP.
Signaling molecule
GTP
GDP
α
γ
β
The α subunit dissociates from the β and γ subunits and triggers a cellular response.
GTP
α
γ
β
Response
GTP is hydrolyzed to GDP.
GTP
P_i
γ
β

Heterotrimeric G proteins have three subunits: *α*, *β*, and *γ*. When a signaling molecule binds to a G-protein-coupled receptor in the plasma membrane, a GDP molecule associated with the *α* subunit is exchanged for GTP. The *β* and *γ* subunits dissociate from the *α* subunit, and a cellular response is triggered either by the *α* subunit or the dissociated *βγ* pair. Hydrolysis of GTP to GDP terminates the signal.

Because of the extra steps, G-protein activated channels and enzymes are slower to activate than their ligand-activated counterparts, but they also remain activated longer. Additionally G-protein-gated channels have more variability as they can be either opened or closed (depending on the channel) by the G-protein subunit.

G-protein-linked receptors have been extensively studied and much has been learned about their roles in maintaining health. Bacteria that are pathogenic to humans can release poisons that interrupt specific G-protein-linked receptor function, leading to illnesses such as pertussis, botulism, and cholera. In cholera ([link]), for example, the water-borne bacterium *Vibrio cholerae* produces a toxin, choleragen, that binds to cells lining the small intestine. The toxin then enters these intestinal cells, where it modifies a G-protein that controls the opening of a chloride channel and causes it to remain continuously active, resulting in large losses of fluids from the body and potentially fatal dehydration as a result.



NOTICE.

PREVENTIVES OF

CHOLERA!

Published by order of the Sanatory Committee, under the sanction of the Medical Counsel.

BE TEMPERATE IN EATING & DRINKING!

Avoid Raw Vegetables and Unripe Fruit !.

Abstain from COLD WATER, when heated, and above all from *Ardent Spirits*, and if habit have rendered them indispensable, take much less than usual.

SLEEP AND CLOTHE WARM !

☞ **DO NOT SLEEP OR SIT IN A DRAUGHT OF AIR.**

Avoid getting Wet !

Attend immediately to all disorders of the Bowels.

TAKE NO MEDICINE WITHOUT ADVICE.

Medicine and Medical Advice can be had by the poor, at all hours of the day and night, by applying at the Station House in each Ward.

CALEB S. WOODHULL, *Mayor.*
JAMES KELLY, *Chairman of Sanatory Committee.*

Transmitted primarily through contaminated drinking water, cholera is a major cause of death in the developing world and in areas where natural disasters interrupt the availability of clean water. The cholera bacterium, *Vibrio cholerae*, creates a toxin that modifies G-protein-mediated cell signaling pathways in the intestines. Modern sanitation eliminates the threat of cholera outbreaks, such as the one that swept through New York City in 1866. This poster from that era shows how, at that time, the way that the disease was transmitted was not understood. (credit: New York City Sanitary Commission)

Section Summary

Initiation of intracellular signal transduction pathways differ depending on the size and polarity of the signal molecule. **Small, non-polar signal molecules** can diffuse easily across the plasma membrane. Therefore their receptor proteins are often (but not always) found in the cytoplasm of the target cell. In most cases these internal receptor proteins are transcription factors that, when activated by the signal molecule, attach to DNA in the nucleus and affect transcription of mRNA. **Larger polar molecules** cannot cross the plasma membrane. Therefore their receptors are part of transmembrane proteins with the active site on the external surface. The signal is transferred to the interior of the cell through the conformation changes that occur when the signal molecule binds to the receptor. In some cases small non-polar signal molecules also bind to cell-surface receptors to initiate their response.

Signal transduction using cell-surface receptors can be classified into four basic types based on the **activation mechanism** and the **mode of action**. The two activation mechanisms are **direct**, where the active site for the signal molecule and the effector protein are on the same protein or a physically linked protein, and **indirect**, where the active receptor site for the signal molecule is on a separate protein from the effector protein. In indirect activation the two proteins communicate using the subunits (α and $\beta\gamma$) of a G-protein. The two modes of action for the effector are either a **channel** or an **enzyme**.

Ligand-gated channels have a receptor and channel on the same transmembrane protein and when the signal molecule attaches the channel opens. **Ligand-activated enzymes** are transmembrane proteins with a receptor domain on the exterior of the plasma membrane and an enzyme domain on the interior. When the signal molecule attaches to the surface receptor the conformation change activates the intracellular enzyme portion. G-protein-linked receptors interact with a G-protein on the cytoplasmic side of the plasma membrane promoting the exchange of bound GDP for GTP. The α and $\beta\gamma$ subunits released by this interaction can interact with enzymes or ion channels on a separate protein to transduce the signal into action. **G-protein-gated channels** respond to an α or $\beta\gamma$ subunit and can cause the channel to either open or close depending on the specific channel activated. **G-protein-activated enzymes** also are activated by an α or $\beta\gamma$ subunit of a G-protein. Both ligand-activated and G-protein-activated enzymes can create **second messengers** that cause further intracellular changes that **amplify and diversify the intracellular response**.

Art Connections

Exercise:

Problem:

[link] HER2 is a receptor tyrosine kinase. In 30 percent of human breast cancers, HER2 is permanently activated, resulting in unregulated cell division. Lapatinib, a drug used to treat breast cancer, inhibits HER2 receptor tyrosine kinase autophosphorylation (the process by which the receptor adds phosphates onto itself), thus reducing tumor growth by 50 percent. Besides autophosphorylation, which of the following steps would be inhibited by Lapatinib?

a. Signaling molecule binding, dimerization, and the downstream cellular response.
b. Dimerization, and the downstream cellular response.
c. The downstream cellular response.
d. Phosphatase activity, dimerization, and the downsteam cellular response.

Solution:

[link] C. The downstream cellular response would be inhibited.

Review Questions

Exercise:

Problem:

What property prevents the ligands of cell-surface receptors from entering the cell?

a. The molecules bind to the extracellular domain.
b. The molecules are hydrophilic and cannot penetrate the hydrophobic interior of the plasma membrane.
c. The molecules are attached to transport proteins that deliver them through the bloodstream to target cells.

d. The ligands are able to penetrate the membrane and directly influence gene expression upon receptor binding.

Solution:

B

Exercise:

Problem:

Why are ion channels necessary to transport ions into or out of a cell?

a. Ions are too large to diffuse through the membrane.
b. Ions are charged particles and cannot diffuse through the hydrophobic interior of the membrane.
c. Ions do not need ion channels to move through the membrane.
d. Ions bind to carrier proteins in the bloodstream, which must be removed before transport into the cell.

Solution:

B

Exercise:

Problem:

Endocrine signals are transmitted more slowly than paracrine signals because ____________.

a. the ligands are transported through the bloodstream and travel greater distances
b. the target and signaling cells are close together
c. the ligands are degraded rapidly
d. the ligands don't bind to carrier proteins during transport

---

Solution:

A

Free Response

Exercise:

Problem:

What is the difference between intracellular signaling and intercellular signaling?

Solution:

Intracellular signaling occurs within a cell, and intercellular signaling occurs between cells.

Exercise:

Problem:

What are the differences between internal receptors and cell-surface receptors?

Solution:

Internal receptors are located inside the cell, and their ligands enter the cell to bind the receptor. The complex formed by the internal receptor and the ligand then enters the nucleus and directly affects protein production by binding to the chromosomal DNA and initiating the making of mRNA that codes for proteins. Cell-surface receptors, however, are embedded in the plasma membrane, and their ligands do not enter the cell. Binding of the ligand to the cell-surface receptor initiates a cell signaling cascade and does not directly influence the making of proteins; however, it may involve the activation of intracellular proteins.

Exercise:

Problem:

Cells grown in the laboratory are mixed with a polar dye molecule that is unable to pass through the plasma membrane. If a ligand is added to the cells, observations show that the dye enters the cells. What mode of action did the ligand initiate?

Solution:

It caused a transmembrane ion channel to open, which allowed the polar dye molecule to move into the cell.

Glossary

autocrine signal
signal that is sent and received by the same or similar nearby cells

cell-surface receptor
cell-surface protein that transmits a signal from the exterior of the cell to the interior, even though the ligand does not enter the cell

chemical synapse
small space between axon terminals and dendrites of nerve cells where neurotransmitters function

endocrine cell
cell that releases ligands involved in endocrine signaling (hormones)

endocrine signal
long-distance signal that is delivered by ligands (hormones) traveling through an organisms circulatory system from the signaling cell to the target cell

enzyme-linked receptor
cell-surface receptor with intracellular domains that are associated with membrane-bound enzymes

extracellular domain
region of a cell-surface receptor that is located on the cell surface

G-protein-linked receptor
cell-surface receptor that activates membrane-bound G-proteins to transmit a signal from the receptor to nearby membrane components

intercellular signaling
communication between cells

internal receptor
(also, intracellular receptor) receptor protein that is located in the cytosol of a cell and binds to ligands that pass through the plasma membrane

intracellular mediator
(also, second messenger) small molecule that transmits signals within a cell

intracellular signaling
communication within cells

ion channel-linked receptor
cell-surface receptor that forms a plasma membrane channel, which opens when a ligand binds to the extracellular domain (ligand-gated channels)

ligand
molecule produced by a signaling cell that binds with a specific receptor, delivering a signal in the process

neurotransmitter
chemical ligand that carries a signal from one nerve cell to the next

paracrine signal
signal between nearby cells that is delivered by ligands traveling in the liquid medium in the space between the cells

receptor
protein in or on a target cell that bind to ligands

signaling cell
cell that releases signal molecules that allow communication with another cell

synaptic signal
chemical signal (neurotransmitter) that travels between nerve cells

target cell
cell that has a receptor for a signal or ligand from a signaling cell